

Think harder about ecstasy

Advocates of therapeutic uses of the drug ecstasy have won the right to research its performance, but opponents continue to snipe. Both sides need to look more deeply into their research agendas.

In a series of studies beginning this year, doctors will give patients with mental problems a drug to try to ease their trauma. But unlike most other drugs tested in clinical trials today, this experimental treatment, MDMA, or ecstasy, is also widely used in clubs around the world. Proponents of the MDMA trials say that they are long overdue, and have been stalled by political concerns (see page 126). Opponents of the trials say that the trials' organizers are playing with fire by introducing potentially toxic substances into perhaps unstable people. So who is right?

The truth is that there is a shortage of scientific logic on both sides of the MDMA debate. Those who demonize the drug are so convinced of its deadly nature that, when Johns Hopkins researcher George Ricaurte reported that even a single dose of MDMA causes debilitating, Parkinson's-like disease in monkeys (G. Ricaurte *et al. Science* **297**, 2260–2263; 2002), the paper was widely touted as proof that MDMA was fatally toxic. Nobody had reported similar findings before, but this mattered little; anti-drug activists had the proof they had long been seeking that MDMA was much worse than just a harmless party drug.

But it later emerged that the findings were fatally flawed because the MDMA had been mixed up with methamphetamine — a drug of abuse already known to be potentially fatal. The available evidence shows that MDMA can cause psychosis, hyperthermia and even death in some people who take the drug recreationally. But there is no research to indicate whether or not this will be a problem in the controlled settings of a clinical trial, or whether such controlled administration might result in long-term brain damage.

On the other side of the debate, MDMA's supporters claim that the drug is a potentially valuable therapeutic aid with little risk of causing lasting brain damage. They argue that studies showing that MDMA leads to cognitive deficiencies are procedurally flawed, and that there is no proof that a few doses of the drug will cause harm. They point to anecdotal reports of the drug's benefits to show that it should be evaluated as a possible aid to psychotherapy for debilitating conditions such as post-traumatic stress disorder. But no rigorously controlled trials have suggested that MDMA has long-term curative properties, and the drug's effects on people with serious mental disturbances have not been evaluated. The drug has also been reported in conjunction with at least 12 cases of psychosis in the medical literature.

Of course, much scientific research is driven by non-scientific motivations, and no drug would ever be proven to work in clinical trials without an interested champion. But in this case, both sides in the debate would benefit by taking a step back and re-focusing their work.

MDMA's detractors in particular would be wise to study the more immediate and realistic side effects of the drug, such as its interactions with other drugs of abuse, and the processes whereby it can sometimes lead to life-threatening hyperthermia. This more reflective approach would help researchers learn more about the origins of and responses to the frightening side effects that sometimes occur when users take MDMA as a recreational drug. This would surely be a better way of ensuring that safe treatments reach the patients who need them. ■

South Africa's new voice

The recent appointment of research minister Mosibudi Mangena bodes well for science, provided that people listen to him.

After a decade of democratic government in South Africa, the face of research remains largely untransformed. Published research has declined significantly in relation to global output, and spending on research and development (R&D) has declined as a proportion of gross domestic product (GDP). White scientists still produce more than 90% of research articles published. Moreover, the research community is ageing, with almost half of authors of research articles now being over 50. Very few able school-leavers are being attracted into research careers and, despite the de-segregation of the country's school system, only a tiny proportion of black scholars leave school with university-entrance qualifications in mathematics and physical science. Those who do are attracted to careers in the professions, which are perceived as more lucrative.

But a glimmer of hope lies in President Thabo Mbeki's choice of minister of science and technology, Mosibudi Mangena, in his new coalition government (see page 117). Mangena has a background as an applied mathematician and has the experience of having served for the past three years as deputy minister of education. In this capacity he has gained an insight, in particular, into the underlying

problem that faces South African science — a lack of qualified teachers and resources to educate its huge school-going population.

Initial indications are that Mangena will not hesitate to articulate solutions, but these will require financing, in the face of many other pressing demands. In his first official appearance as minister last week — launching National Science Week — Mangena made it clear that he regards reform of the science education system as a priority. He has been no less forthright in advocating increased state spending on R&D. The question is whether this will cut any ice with government, particularly as, like his predecessor, he represents a minority party. But the South African cabinet would do well to give him a good hearing.

There are also signs that the government's approach to AIDS is becoming more pragmatic, at least in its election promises. By being equally forthright in this sphere, Mangena could contribute to healing the rift between government and scientists that emerged four years ago following Mbeki's rejection of the prevailing understanding of the disease. Regrettably, Mbeki has not replaced his health minister, Manto Tshabalala-Msimang, whose tardiness in executing changes in anti-retroviral policy remains an obstacle to progress. ■



Secret agent

Text software fills in the blanks on official documents
p116

**Up in the air**

London medical institute faces uncertain future
p117

**Under pressure**

NIH considers rule changes to halt conflicts of interest
p119

**Net gain**

Fishermen trawl for ideas to protect wildlife
p120

Iraqi killings prompt calls for US to evacuate weapons scientists

Jim Giles, London

The assassination of several of Iraq's former weapons scientists has hit US plans to employ them to help rebuild the war-torn country. The killings, together with the deteriorating security situation in Iraq, have led some non-proliferation experts to call for the researchers to be evacuated from the country.

Between five and ten scientists have been killed in the past six months, according to a US Department of State official who runs programmes aimed at keeping former weapons scientists in employment. "The most common explanation is that they've shown an interest in working with the coalition," says the official, who declined to be identified by name and who returned from Iraq earlier this month.

Between them, the Iraqi scientists hold considerable knowledge of chemical, biological and nuclear weapons from programmes that now seem to have been defunct long before the US-led coalition invaded Iraq in March 2003. But the killings are only the latest setback in plans to redirect their knowledge and skills. Non-proliferation experts who wanted to work with Iraqi scientists were angered when initial responsibility for contacting them was given to military forces. Some scientists hid, fearing that they could be taken prisoner (see *Nature* 423, 371; 2003).

Such independent experts have since left Iraq because of security concerns, further weakening non-proliferation efforts. And David Albright, a former nuclear-weapons inspector in Iraq, says these problems mean that attempts to keep researchers in Iraq should no longer be a priority for the US government. "They should shift the programme to getting people out," says Albright, who now heads the Institute for Science and International Security in Washington. "There are scientists with secret documents who could go to Iran or Syria."

Such a change in policy would come too late for Majid Hussein Ali, a nuclear scientist reported to be at Baghdad University. Ali was not directly involved in weapons research, but he was said to have met with US weapons inspectors. He was killed by an unknown gunman in Baghdad in February.



Shattered: the fall-out from the Iraq conflict is putting scientists' lives in danger.

Despite the death of Ali and other researchers, state-department officials insist that most scientists want to stay in their country. Officials have visited Iraq regularly this year, and say that they were able to win the confidence of Iraqi scientists by distancing themselves from the military activities of the coalition forces.

Job creation

The state department sought to ramp up its activities last November with a US\$2-million programme aimed at identifying former weapons researchers and finding them work in Iraq (see *Nature* 426, 371; 2003). Since then, officials have drawn up a list of 400 scientists, engineers and technicians who had worked on weapons research and related fields. Officials say that about 75 of these people are unaccounted for, but nearly all of the others have been located in Iraq.

The officials add that these researchers would stay in Iraq if meaningful work can be found for them. Most are currently employed in industry and academia, at least in theory. But many universities and other facilities have been closed by the invasion and subsequent insurgencies.

"They are all employed in the sense that they get a pay cheque," says the state-depart-

ment official. "But some are very unhappy because they have nothing to do." The official is trying to raise \$40 million for reconstruction projects over the next three years. "We're talking to coalition partners now," he says.

State-department staff have meanwhile established an International Center for Science and Industry in Baghdad, consisting of office space that they say will be used to house Iraqi researchers who will determine how any reconstruction money will be spent.

Many Iraqi scientists have criticized schemes by outsiders to unite the country's researchers, officials at the state department acknowledge. They say that scientists felt excluded from an attempt by a largely expatriate group of Iraqi researchers to form an Iraqi academy of science (see *Nature* 426, 484; 2003). By ensuring that local scientists play a prominent role in the new centre, the officials hope that the facility will be accepted as legitimate by Iraqi researchers.

But Albright, who initially backed the state department's programme, is worried about what will happen to the former weapons scientists. He went to Iraq last year and helped US officials to locate many of them. Such trips ended in the autumn, as the security situation worsened. Even then, Albright says, "everyone wanted to get out".

KHALID MOHAMMED/AP

Bush pressured as Nancy Reagan pleads for stem-cell research

Erika Check, Washington

Almost three years after President Bush laid down a policy restricting the use of public funds in embryonic stem-cell research, calls are growing for the White House to revisit the rules.

On 8 May, Nancy Reagan, former first lady and an icon of Bush's Republican party, spoke publicly for the first time of her support for stem-cell research. She had written letters in favour of it before but her speech, at a benefit dinner in Los Angeles for the Juvenile Diabetes Research Foundation, is seen by supporters of the research as a significant public-relations breakthrough.

Reagan said that she had been moved to support research using stem cells through watching her husband succumb to Alzheimer's disease. "Ronnie's long journey has finally taken him to a place where I can no longer reach him," she said. "We cannot share the wonderful memories of our 52 years together, and I think that is the hardest part. I am determined to do whatever I can to save other families from this pain."

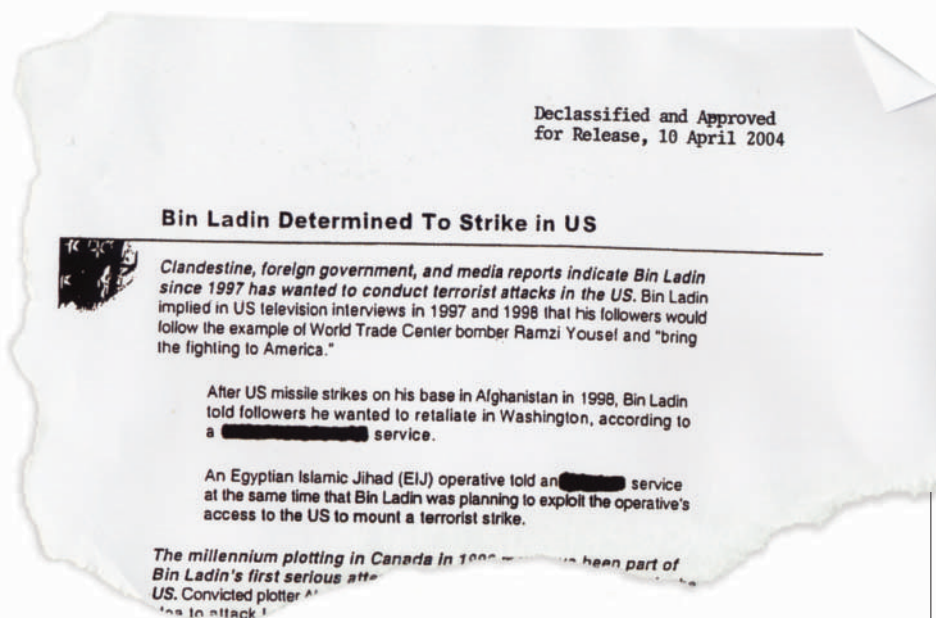
A few days earlier, on 4 May, the majority leader in the Senate, Bill Frist (Republican, Tennessee), said he thought the time had come to review President Bush's policy. The rules let researchers use public funds to work on embryonic stem-cell lines only if the lines were derived before the day the policy was announced — 9 August 2001.

"Momentum is building in the research done, and in Congress," says a Republican staff member for the Senate Committee on Appropriations.

Last month, 206 members of Congress, including 36 Republicans, sent a letter to President Bush asking him to expand his policy. Among its signatories were members of Congress who had previously opposed the research, such as lawmaker Dana Rohrabacher (Republican, California). Rohrabacher told reporters that he had changed his mind after hearing from patients who hope the research will help cure their diseases.

A similar letter is circulating in the Senate, and White House officials have indicated that the president will meet with the authors of the House letter.

Although the National Institutes of Health estimated that researchers would be able to work on 78 stem-cell lines, fewer than 20 are actually available today. And biologists have raised doubts about the suitability of these for clinical research. ■



US intelligence exposed as student decodes Iraq memo

Declan Butler

Armed with little more than an electronic dictionary and text-analysis software, Claire Whelan, a graduate student in computer science at Dublin City University in Ireland, has managed to decrypt words that had been blotted out from declassified documents to protect intelligence sources.

She and one of her PhD supervisors, David Naccache, a cryptographer with Gemplus, which manufactures banking and security cards, tackled two high-profile documents. One was a memo to US President George Bush that had been declassified in April for an inquiry into the 11 September 2001 terrorist attacks. The other was a US Department of Defense memo about who helped Iraq to 'militarize' civilian Hughes helicopters.

It all started when Naccache saw the Bush memo on television over Easter. "I was bored, and I was looking for challenges for Claire to solve. She's a wild problem solver, so I thought that with this one I'd get peace for a week," Naccache says. Whelan produced a solution in slightly less than that.

Demasking blotted out words was easy, Naccache told *Nature*. "Optical recognition easily identified the font type — in this case Arial — and its size," he says. "Knowing this, you can estimate the size of the word behind the blot. Then you just take every word in the dictionary and calculate whether or not, in that font, it is the right size to fit in the space, plus or minus 3 pixels."

A computerized dictionary search yielded 1,530 candidates for a blotted out word in this sentence of the Bush memo: "An Egyptian Islamic Jihad (EIJ) operative told an [redacted] service at the same time that Bin

Ladin was planning to exploit the operative's access to the US to mount a terrorist strike." A grammatical analyser yielded just 346 of these that would make sense in English.

A cursory human scan of the 346 removed unlikely contenders such as acetose, leaving just seven possibilities: Ugandan, Ukrainian, Egyptian, uninvited, incursive, indebted and unofficial. Egyptian seems most likely, says Naccache. A similar analysis of the defence department's memo identified South Korea as the most likely anonymous supplier of helicopter knowledge to Iraq.

Intelligence experts say the technique is cause for concern, and that they may think about changing procedures. One expert adds that rumour-mongering on probable fits might engender as much confusion and damage as just releasing the full, unadulterated text.

Naccache accepts the criticism that although the technique works reasonably well on single words, the number of candidates for more than two or three consecutively blotted out words would severely limit it. Many declassified documents contain whole paragraphs blotted out. "That's impossible to tackle," he says, adding that, "the most important conclusion of this work is that censoring text by blotting out words and re-scanning is not a secure practice".

Naccache and Whelan presented their results at Eurocrypt 2004, a meeting of security researchers held in Interlaken, Switzerland, in early May. They did not present at the formal sessions, but at a Tuesday evening informal 'rump session', where participants discuss work in progress. "We came away with the prize for the best rump-session talk — a huge cow-bell," says Naccache. ■

South Africa names head of science ministry

Michael Cherry, Cape Town

An applied mathematician has been appointed minister of science and technology in the new South African government — and has pledged to start bringing more young people into science.

Since apartheid collapsed ten years ago, the average age of South Africa's scientific workforce has gone up, with few young researchers being recruited. The country's share of global published scientific output fell from 0.8% in 1990 to 0.5% in 2000 (A. Pouris *S. Afr. J. Sci.* **99**, 425–427; 2003).

But the new minister, 56-year-old Mosibudi Mangena, who holds a master's degree from the University of South Africa, has said he will confront the problem. "To redress this situation, we urgently need to identify talent and nurture it," he said shortly after his appointment on 29 April.

Mangena, a former political prisoner and exile, is the head of a tiny minority left-wing party, the Azanian People's Organisation, which secured only one seat in parliament in the recent elections. He was previously deputy minister of education. But because



Mosibudi Mangena plans steps to woo young scientists.

he is not a member of President Thabo Mbeki's African National Congress, which dominates the new government, some scientists fear that Mangena will have a tough job making his voice heard.

He says that his main task will be to persuade the cabinet to reverse the past decade's 30% decline in South African spending, as a proportion of the economy, on research and development. This has

recently increased slightly, from 0.69% in 1997–98 to 0.76% in 2001–02, but it is still far from the target of 1% by next year that was set out in a national research strategy published in 2002.

Mangena says he is confident that the government will see the wisdom of increasing research spending, despite other pressing demands for social services. "The president has already committed the government to a higher profile for science and technology," he says. "Once we have committed more funds, we shall have to find ways of persuading the private sector to follow suit."

But the biggest problem facing the new minister is to attract more young people

into science, which is widely viewed as arcane and poorly paid. Only 4,800 black pupils achieved a higher-grade pass in mathematics in the school-leaving examinations last year, and science courses were the least popular at universities.

The apartheid government, "for all its faults, provided resources for research", concedes Mangena. As a consequence, most of the country's research base is concentrated in its science councils and in a few historically predominantly white universities. "I don't think that it is feasible for all the country's universities to be doing top-level research," he says, "but it is essential that research institutions change radically, and do not remain a white preserve."

Khotso Mokhele, president of the National Research Foundation, the country's main grant funding agency, says he was "delighted" with Mangena's appointment because of his qualifications as an applied mathematician and also his understanding of the problems facing researchers.

Fritz Hahne, director of the African Institute for Mathematical Sciences in Cape Town, also welcomed Mangena's appointment "on account of his strong track record of support for science, and mathematics in particular".

Researchers fear break-up of UK medical institute

Laura Nelson, London

Scientists at the National Institute for Medical Research (NIMR) at Mill Hill in north London are worried that their institute could soon be split up.

The Medical Research Council (MRC) has confirmed that it is revisiting a previous decision to keep the NIMR — one of Britain's premier centres for basic medical research — on one site.

The MRC has been trying to shift the emphasis of the institute more towards clinical research, and last year suggested that the site should be moved closer to a research hospital, proposing Addenbrooke's Hospital in Cambridge (see *Nature* **423**, 573; 2003).

After protests by the institute's researchers, a task force decided in March this year that the institute would not move to Cambridge, but would instead stay put or move to a different, single site in London. It rejected another proposal that the institute be broken into pieces to be closer to hospitals around London.



All together: researchers wrestle with blood plasma shortly after the National Institute of Medical Research opened in 1950.

But last month the task force had a change of heart. In response to an MRC-commissioned study on investment requirements, carried out by consultancy company Ove Arup and Partners, it reinstated the option of "an institute located on more than one site".

The task force has since consulted London hospitals and colleges that could potentially offer sites to the new NIMR. "We asked what they could do and they have offered a variety of proposals," says Colin Blakemore, chief executive of the MRC and chairman of the task force reviewing the institute's fate. "We would like to keep the NIMR as one institute. But there are practical constraints," he says. "We have to look at more modest possibilities."

Some of the roughly 600 NIMR scientists, students and staff are now convinced that their institute is destined to be split into pieces around London. "The idea of them finding a single site is unlikely," says Barrie Brown, head of labour

relations for Amicus MSF, one of the NIMR's trade unions. "Our priority is to retain a single-site institute, because we think it is a driving force for medical research," says Brown.

The task force aims to present a London site proposal to the MRC in July.

Scientists complain government cash is no rise in real terms

Carina Dennis, Sydney

The Australian government has announced a A\$5.3-billion (US\$3.7-billion), seven-year investment package aimed at boosting science and innovation.

The plan, which was launched by John Howard, the prime minister, on 6 May at Parliament House in Canberra, includes extra funding for the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the nation's largest research agency.

Howard called it "the biggest ever commitment by a government in the area of science and innovation".

But the big bucks in the package — A\$1 billion — go to a new entity that will try to nurture industrial innovation. Some scientists, together with members of the opposition Labor party, say that the research part of the package is inadequate.

The CSIRO will receive an increase of A\$305 million dollars over seven years, on top of its existing annual budget of some A\$570 million. "The increase will not keep up with the growth in the economy," complains Snow Barlow, a plant biologist at the University of Melbourne and president of the Federation of Australian Scientific and Technological Societies.

"At least it provides some planning continuity for agencies and programmes," says Ken Baldwin, a physicist at the Australian National University in Canberra.

As part of the package, the Australian Research Council, which provides grants for basic research at universities, will complete a five-year plan to double its budget to about A\$560 million by 2006, and keep it at that level until 2011. The proposal also includes A\$540 million to aid research collaborations and A\$200 million for independent medical institutions. ■



John Howard: the Australian prime minister earmarks funds for industrial innovation.

Icelandic database shelved as court judges privacy in peril

Alison Abbott, Munich

Iceland's supreme court has ruled that the transfer of a dead patient's health data to a proposed genetic database would infringe the privacy rights of the man's descendants.

The ruling — which was made late last year but published in English only last month — casts further doubt over the nation's plans for a Health Sector Database to hold centralized electronic health records on its population.

The plans attracted global attention when they were formulated in 1998 (see *Nature* 396, 395; 1998). They would use Iceland's population, claimed by some to be unusually homogeneous, to pioneer genetic population studies.

But the company contracted to build the database — deCODE Genetics of Reykjavik — has already postponed its development. The plans were quietly put on ice in 2002 after the company was unable to reach agreements with regulators about what information the database would contain or with hospitals about who would pay for it.

On 27 November, the court found in favour of Ragnhildur Gudmundsdottir, an 18-year-old student, who did not want her dead father's health records to be transferred to the database if plans went ahead. The court said that including the records in the database might allow her to be identified as an individual at risk of any heritable disease her father might be found to have had — even though the data would be made anonymous and encrypted.

The possibility of such identification is increased by the fact that the Health Sector Database would allow information to be linked with data from other genetic and genealogical databases, the court found.

The ruling has been interpreted to mean that the 1998 law governing the creation of the database is unconstitutional because it fails to protect personal privacy adequately.

But the construction of the database is already in doubt. DeCODE, which in 1999 was granted an exclusive 12-year licence to build it, was unable to agree with the Icelandic Data Protection Authority or Iceland's National Bioethics Committee about use of the data. To find genes associated with disease, the company had wanted to combine individuals' health data with genomic data from blood samples, which it was gathering from the community, and a genealogical database it has created. The company was also unable to reach agreement on who should pay for the project.

Sigrun Johannesdottir, head of the Data Protection Authority, says that it declined to



Fellow feeling: many Icelanders opted out of a database that would be used for genetic studies.

grant deCODE the free access that it sought. European Union rules, she says, demand that only averaged data can be extracted from it, not the individual data, which are more scientifically valuable.

Magnus Petursson, director of the University Hospital in Reykjavik, which serves more than half of the population, says that the hospital began work on the project in 2000 but negotiations petered out. "We have not heard whether they will start again," he says. Kári Stefánsson, chief executive of deCODE, says that the company is "still working with the government to put together the database".

David Gunnarsson, secretary of Iceland's health ministry, says that the government is still interested and has been closely following the project's progress. He denies that the court ruling undercuts the 1998 act. "The first issue for the government," Gunnarsson says, "is to find a way to resolve the discussions according to what the law allows, and to everyone's satisfaction".

Edward Farmer, a spokesman for deCODE, says that the company has already collected blood samples from a total of 110,000 adults in Iceland — more than half the adult population — during 50 disease-related genetic studies. So far, deCODE is not permitted to use the samples collected in work on unrelated diseases.

The database plan was controversial in Iceland: more than 20,000 people actively opted out of it. Now long-term critic Skúli Sigurdsson, a science historian affiliated to the University of Iceland, predicts that the project will not happen "given the complexity of data mining and personal privacy issues, and the tremendous inertia of health politics". ■

Additional reporting by Jim Giles, London.

NIH urged to rewrite rules on consultancies

Erika Check, Washington

The National Institutes of Health (NIH) looks set to change its employment rules to address criticism about conflicts of interest among its senior staff.

The changes were proposed on 6 May by an outside panel appointed by Elias Zerhouni, the NIH's director, after an article in the *Los Angeles Times* in December exposed shortcomings in the agency's existing conflict-of-interest policies.

The panel, which was chaired by Bruce Alberts, president of the National Academy of Sciences, and Norman Augustine, chairman of the defence contractor Lockheed Martin, said that the NIH should bar its most senior officials from earning money by consulting for industry or academic institutions.

"The NIH has honestly sought to apply the laws as it understands them, but there is room for substantial improvement in its conflict-of-interest policies," Augustine says.

Scientists who are allowed to do consultancy work should limit their compensation to 50% of their government salaries, the panel said, and spend no more than 400 hours a year on the work. Clinicians at the NIH would be subject to slightly less restrictive rules. But nobody at the NIH should be allowed to accept stocks or stock options as payment, it recommended.



Elias Zerhouni: considering fundamental changes to prevent conflicts of interest.

To allay any negative impact on the agency's ability to recruit staff, the panel suggested that the health secretary raise the cap on the salaries that the NIH can pay them.

Zerhouni says that he will examine the

recommendations and consult with his advisory committees before implementing any changes. "The recommendations are very clear, and they would profoundly change the way you would manage the agency," he says. But some of them, such as the salary changes, would require action from Congress.

Scientists' groups have supported the panel's findings. "I believe that what they came up with will provide more adequate protection without decreasing the attractiveness of the agency to scientists," says Jordan Cohen, president of the Association of American Medical Colleges.

"I think the report goes in the right direction," says Harold Varmus, former director of the NIH, who changed the agency's rules to let its scientists work with industry. But he adds that the strongest restrictions should apply to only a few scientists.

Varmus is critical of the the proposed bar on stock options, however. "I don't agree that having any stock — especially in a company that doesn't have publicly traded stock and may be worth nothing — is necessarily a conflict of commitment," he says.

But some patient advocates argue that the proposed rules don't go far enough. "They only partially address the underlying, corrosive conflicts of interest," says Vera Sharav, president of the New York-based Alliance for Human Research Protection.

R. L. WOLLENBERG/UPI PHOTO

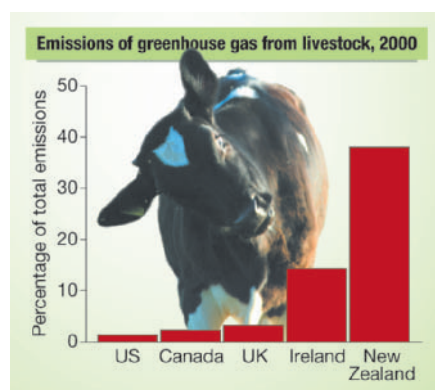
Vaccine targets gut reaction to calm livestock wind

Carina Dennis, Sydney

The belches of millions of cows and sheep may not immediately seem like a serious problem. But for some countries they send greenhouse-gas emissions through the barnyard roof, thanks to methane-producing bacteria in the animals' guts.

Scientists at Australia's largest research body, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), now hope to change this with a vaccine that targets methane-producing bacteria. In a test on 30 sheep, the team's vaccine reduced methane emissions by 7.7% (A. D. G. Wright *et al. Vaccine* doi:10.1016/j.vaccine.2004.03.053; 2004).

Making methane is a ruminant's way of letting off steam. As food ferments in the animal's first stomach, or rumen, hydrogen is produced and reacts with carbon to form methane, which is then exhaled. Although this accounts for less than 3% of total greenhouse-gas emissions in fossil-fuel burning countries such as Britain and the United States, it makes up nearly 40% of emissions in agricultural New Zealand (see chart).



Andre-Denis Wright a molecular biologist at CSIRO Livestock Industries in Perth, who led the study, says that the team's vaccine should cut emissions by 20% once it matches the microbes in sheep better — rumen bacteria vary from region to region and from season to season.

But critics point out that this diversity will also limit the vaccine's usefulness. "There are a huge number of microbes in the rumen that can't be cultivated. And if you can't cultivate them, you can't raise

antibodies and can't make a vaccine against them," says Athol Klieve, a rumen microbiologist from the Queensland Department of Primary Industries and Fisheries in Brisbane.

Killing off some methane-producing bacteria in the gut may also open the way for other methane-producing microbes to take their place.

Wright's team hopes to get round these problems by tailoring the vaccine to target a protein common to many methane-producing bacteria. Other groups are working on different solutions. Athol, for example, plans to introduce bacteria from the guts of kangaroos into cows in the hope that these will out-compete methane-producing microbes. The kangaroo bacteria produce acetate instead.

But the vaccine will never eliminate methane production altogether, says Roger Hegarty, a livestock researcher for the New South Wales state government's agriculture department. Complete blockage of burping could result in an unhealthy build-up of hydrogen in the animals, he says.

FREEFOTO.COM

Disarmament dispute leaves nuclear treaty's future in doubt

Paris A question mark is hanging over the future of the Treaty on the Non-Proliferation of Nuclear Weapons, after a meeting to set the agenda for an important review of the treaty ended without agreement on 7 May.

Under the treaty, known as the NPT, the 'weapons states' — China, the United States, France, Britain and Russia — agree to supply civil nuclear technology to countries in return for them not developing weapons. The weapons states also commit to disarming their own nuclear weapons.

But after last week's meeting in New York, the non-nuclear states claimed that the weapons states want to avoid the issue of disarmament at next year's conference. "The foundation deal of the treaty is being threatened," says Rebecca Johnson, director of the London-based Acronym Institute for Disarmament Diplomacy, which promotes nuclear disarmament.

The United States also claimed that the NPT failed to stop countries such as Iran from developing weapons. It wants stricter measures to limit nuclear reprocessing plants, and to deal with countries that divert nuclear capabilities to military use.



Trawl aboard: commercial fishermen are being asked to float ideas for better equipment.

Contest casts net for ideas to save marine wildlife

San Diego Fishermen have been offered a US\$25,000 lure to develop commercial fishing gear that helps to prevent the accidental killing of seabirds, sea turtles and marine mammals.

With tens of thousands of species accidentally harmed each year by fishing, a coalition of environmental and industry groups has created the International Smart Gear Competition to help reduce the problem. The competition was announced on 3 May at the 4th World Fisheries Congress in Vancouver, Canada.

Fishermen already use acoustic alarms to steer marine mammals away from nets, and circular hooks that reduce turtle catches (see *Nature* 428, 594; 2004). Other organizations, such as BirdLife International, also host prizes for such inventions. Last month, two fishermen from New Zealand and Australia netted €18,000 (US\$21,000) for their concoction of fish-liver grease that deters albatrosses from frequenting waters where long-lined boats are fishing.

♦ <http://smartgear.org>

Charity worker bids to play animal name game

New York Next time you stumble across a new species, think twice about naming it after yourself — you might be able to raise some cash by selling the privilege instead.

Last week, a New York-based conservation charity, the Wildlife Trust, auctioned off the opportunity to christen a single-celled microbe. The organism, which lives in the gut of the endangered gecko *Phelsuma guimbeaui* in Mauritius, was discovered by a team from the charity.

Bidding at the 3 May fundraising bash was fierce. As the hammer fell, Cynthia Stebbins, one of the charity's two vice-presidents, scooped the opportunity to name the beast for just US\$2,600. "How

often to do you get to do such a thing?" she says. She plans to name it *Eimeria stebbinsi* after the microbe's genus and her married name.

US seeks Midas touch to model bioterror attacks

New York The US National Institute of General Medical Sciences is seeking to harness computer models to predict the effects of disease epidemics and bioterror attacks.

On 4 May, the institute announced the first four recipients of grants under the project, which is known as MIDAS, or Models of Infectious Disease Agent Study. The recipients will share US\$28 million over five years to build models that incorporate detailed information about the location and behaviour of individuals in a city before and during an outbreak or an attack. The institute hopes that these will aid the design of better strategies to deal with such events.

Stephen Eubank of the Los Alamos National Laboratory in New Mexico, one of the recipients of the four grants, has already built one model. He and his colleagues worked out how smallpox is likely to spread among the 1.5 million people in Portland, Oregon (see page 180). Eubank

Acid test shows microbes carved caverns

London Some of the world's largest and most spectacular caves were created by some very small builders.

Researchers have long thought that acids created by chemical reactions between thermal springs and air etch away at rocks to form karst landscapes such as the Carlsbad Caverns in New Mexico (pictured) and the Frasassi caves in Italy.

But Annette Summers Engel and her colleagues at the University of Texas at Austin have discovered an alternative explanation.

They found that Lower Kane Cave in Wyoming — long considered a classic example of a cave formed by this mechanism — was instead eaten away by microbes (A. S. Engel, L. A. Stern and P. C. Bennett *Geology* **32**, 369–372; 2004).



The microbes devoured hydrogen sulphide from the thermal springs below, the team found, producing sulphuric acid. This converts limestone to gypsum, which dissolves more easily in water.

now hopes to use the MIDAS funding to extend this model to cover the larger city of Chicago.

Engineer takes control at Italy's research body

Rome After a year of government-directed reform, the CNR, Italy's main basic-research organization, has a new president.

The Italian government last year put the CNR in the hands of a commissioner

who was charged with restructuring the organization to make its 100 or so research institutes more efficient.

Fabio Pistella, an engineer at the University of Rome, has now been appointed to run the reformed CNR. Pistella is best known for his administrative skills. He is a former director-general of ENEA, Italy's energy and environment agency, and has also served as director of the National Institute for Applied Optics in Florence.

Time lords

Geologists have come to an international agreement about the timeline of Earth's history. But the results are not quite set in stone, as John Whitfield discovers.

Silurian, Devonian, Triassic: the names seem as solid and permanent as rocks themselves. But in fact, like fashions in hair or hem-length, the geological divisions of our planet's timeline are prone to change. Geologists argue over what defines the boundaries between these periods, and disagree over exactly how old they are. The end of the Jurassic period, for instance, has wobbled by more than 30 million years since it was dated in the 1930s (refs 1, 2).

This summer, the first fully revised geological timescale for 15 years will be published³, helping to clear up some of the confusion. The new scale will help Earth scientists speak to one another in a common language, and the dates will help to answer questions about what caused various mass extinctions and changes in climate. The project will even have an impact on oil companies, which base their exploration programmes in part on estimates of the age of rocks.

The timescale (see below) has been a long time in the making. "I thought it would take two years," says one of its editors, James Ogg of Purdue University in West Lafayette, Indiana. Instead it took seven. Ogg is also secretary-general of the International Commission on Stratigraphy (ICS), the organization that coordinated the project.

But despite being the most complete work on dates so far — and the first to be endorsed by the ICS's parent body, the



Rock of ages: geologists are using Australia's Flinders Ranges to define a Precambrian period.

International Union of Geological Sciences (IUGS) — the new timescale is not definitive. Its assembly has rekindled a long-standing debate on how to define the earliest geological time periods, and many of those boundaries look set to change in the next few years. And as the dates are re-examined, the timescale may even widen to encompass rocks beyond Earth.

Event horizons

Over the past 150 years, geologists have struggled to unravel Earth's history. To a large extent, they have relied on significant events, such as the appearance of a specific fossil, or a reversal in the planet's magnetic field, to define the boundary between two time periods. Having defined these physical boundaries, researchers then attempted to date them. But geologists in different parts of the world used different rocks as benchmarks, leading to disagreements over the exact definition of each period.

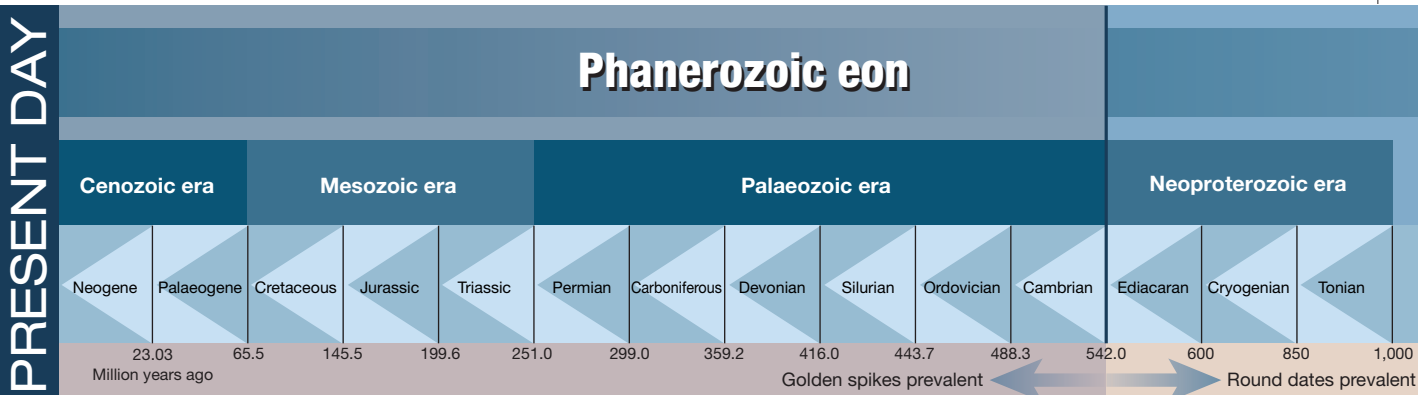
To resolve the issue, stratigraphers are deploying 'golden spikes' — also known as global standard stratotype-section and points

(GSSPs). These are locations where a good example of a worldwide event can be found, nominated by working groups within the ICS and then ratified by the IUGS. Once a spike is set, that rock remains the boundary of a time period, even if estimates of its age change.

The first GSSP was established in 1972 near the town of Klonk, in what is now the Czech Republic. Here can be found pristine fossils of an extinct marine invertebrate called a graptolite, the appearance of which divides the Silurian from the Devonian⁴. But the spike system was slow to catch on. For the 1989 timescale, fewer than 15 out of the 91 geological stage boundaries between now and the beginning of the Cambrian period had been marked with golden spikes.

In 1999, the IUGS executive called for a concerted effort to find GSSPs⁵, and the past few years have seen a surge of activity. Political as well as scientific arguments ensued. In the debate over where to put the spike separating the Permian from the Triassic, the relative inaccessibility of sites in Kashmir and Iran helped to rule them out in favour of one in Zhejiang Province, China. For Chinese

B. EDMAN/ISTOCK



stratigraphers, this was the equivalent of winning a bid to host the Olympics. They hoisted a six-metre-tall monument on the spot to commemorate the achievement.

The spike count now stands at 50. But finding GSSPs is only half the battle — the divisions need dates too. The new timescale takes advantage of a growing ability to date rocks using astronomical events, says Felix Gradstein of the University of Oslo, Norway, ICS chairman and another of the timescale's editors. Using this technique, the past 23 million years — the Neogene period — has been dated to within plus or minus 40,000 years³. "This is the single most exciting scientific development in the new timescale," says Gradstein. "Every geologist should be amazed by this."

The dating game

Astronomical dating back-calculates cycles in Earth's orbit to work out its climate. This is then matched to climate records in the rocks: variations in the ratios of different oxygen isotopes reflect the size of the polar ice caps at the time the rock was laid down, and calcium carbonate content reflects the sea's temperature, for example⁶. The chaotic nature of orbital cycles means that, for now, this method is accurate only to about 65 million years ago. But researchers are confident that it will soon extend back 150 million years.

To pin down the age of older rocks, geologists rely on radiometric dating, which tracks the radioactive decay of elements within a sample. But in the past decade, it has become clear that the results from different techniques and different labs don't agree.

The changing age of the Jurassic is a case in point. In 1987, the period was estimated to have ended 131 million years ago, based on the amount of potassium that had been converted into argon in a mineral called glauconite². But it was later discovered that argon seeps out of glauconite, making the mineral seem younger than it actually is. The new timescale used potassium-argon dating of basalt to put the end of the Jurassic at 145.5 million years ago.

"Most people will tell you that a measurement more than five years old is obsolete,"

says Gradstein. The inconsistencies highlighted by the timescale's construction prompted researchers to propose an international network of laboratories, all using a standard procedure, to which anyone could send a rock sample for dating⁷. These labs should begin publishing results in about a year.

Uncertainty over dates is a particular problem for the early life of Earth. When these time periods were established, it was thought that there would be too few benchmarks to define GSSPs: rocks get

rarer as they get older, and there are no large fossils from that long ago. Instead, each period is anchored by round dates rather than golden spikes. Most of the Precambrian, the first four billion years of Earth's history, is chopped up in this way.

Some Earth scientists have railed against this system ever since the ICS decided on it in the 1980s, complaining that it is inconsistent. "There's been a ferocious debate between a

"There's been a ferocious debate between a bunch of field-based mavericks and the committees"

bunch of field-based mavericks and the committees," says geologist Euan Nisbet of Royal Holloway, a college of the University of London in Egham.

It now looks as if the argument is going the mavericks' way. In March this year, the first golden spike was driven into Precambrian time — rocks in Australia's Flinders Ranges that record the end of glaciations thought to have covered Earth some 600 million years ago have become the starting point for the Ediacaran.

More ancient spikes are set to come. The new chairman of the ICS Precambrian sub-commission, Wouter Bleeker of the Geological Survey of Canada, Ottawa, is on a mission to use rock features to divide up the deep past. Ultimately, Bleeker hopes that the geological record can be extended through space to cover the entire Solar System, giving geologists and planetary scientists a



Geologists stand by the Chinese monument that marks a 'golden spike' between two ages.

common language. The Moon's birth, in a collision between Earth and another planet-sized body about 4.5 billion years ago, would be a good candidate for a spike, he argues, marking the beginning of the Hadean eon — a name that is not yet official, but which many geologists use to describe the time before the Archaean. Using this as a spike would synchronize the terrestrial timescale with the lunar, which currently has its own periods. "On the Moon, there are many obvious stratigraphic boundaries that

probably have relevance on Earth," says Bleeker. "It's very important to try to synchronize where we can."

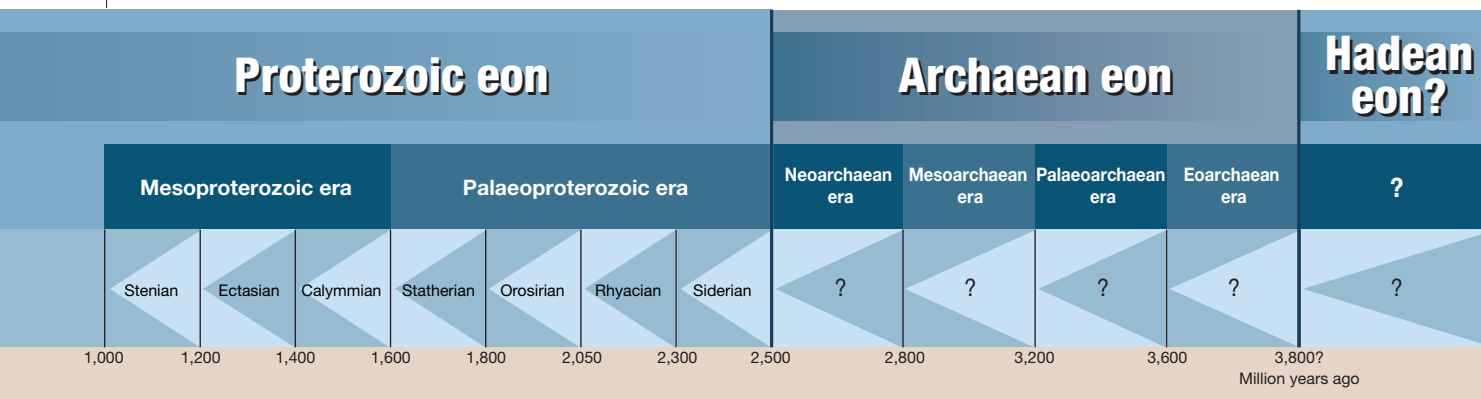
To define such ancient spikes, geologists will need to relax their criteria. There is no worldwide record of the Ediacaran changes seen in Australia, and this will be true for other spikes in the distant past. Nevertheless, Bleeker hopes to have a broad framework of Precambrian spikes worked out by 2008. "We're probably going to move towards using the rock record to divide the Precambrian," agrees Ogg.

By 2008, all of the geological time divisions in the past 600 million years should also bear golden spikes. Good news for geologists seeking to speak a common language; but more work for Gradstein and his team. "We should have a second edition then," he says.

John Whitfield is a science writer based in London.

1. Holmes, A. *The Age of The Earth* (Nelson, London, 1937).
2. Haq, B. U., Hardenbol, J. & Vail, P. R. *Science* **235**, 1156–1167 (1987).
3. Gradstein, F. M., Ogg, J. G. & Smith, A. G. (eds) *A Geologic Time Scale 2004* (Cambridge Univ. Press, in the press).
4. Chlupáč, I. & Kukal, Z. in *The Silurian–Devonian Boundary* (ed. Martinsson, A.) IUGS Ser. A, Vol. 5, 96–109 (IUGS, Stuttgart, 1977).
5. Vai, G. B. *Episodes* **24**, 29–31 (2001).
6. Hilgen, F. J., Krijgsman, W., Langereis, C. G. & Lourens, L. J. *Eos* **78**, 285, 288–289 (1997).
7. Clarke, T. *Nature* **425**, 550–551 (2003).

♦ www.stratigraphy.org



The ups and downs of ecstasy

The clubbers' drug is now being studied as a treatment for anxiety and post-traumatic stress disorder. Erika Check charts its rocky road from the psychedelic underground to the psychiatric clinic.

On 16 April, a patient suffering from post-traumatic stress disorder visited a psychiatrist's office in Charleston, South Carolina, to try a new experimental treatment. She is the first of 20 people who have signed up for the trial. Like the others, she suffers from flashbacks, nightmares and years of painful memories, despite receiving counselling and antidepressant drugs. Each of the volunteers will undergo further counselling, this time in combination with doses of 3,4-methylenedioxymethamphetamine, or MDMA — the active ingredient in the recreational drug ecstasy.

The study is a landmark victory for a group that has fought long and hard to bring drugs such as MDMA into the clinic. The Multidisciplinary Association for Psychedelic Studies (MAPS), based in Sarasota, Florida, has been arguing since the mid-1980s that MDMA is a powerful therapeutic tool. Its leader, Rick Doblin, now hopes that a similar study in Spain will be allowed to resume after a two-year hiatus. He is also working with doctors in Israel to plan trials there. And MAPS has donated money to Harvard Medical School in Boston, where psychiatrist John Halpern is planning a trial to test whether MDMA can relieve anxiety and pain in end-stage cancer patients.

MDMA is classified as a hallucinogen, and other members of this class of drugs are also attracting renewed interest as treatments for psychiatric conditions (see 'Can mushrooms work their magic?', overleaf). The primary effect of these drugs is to alter perception, cognition or mood¹. People who take ecstasy recreationally generally experience a feeling of euphoria. And in the 1980s, a few patients started taking the drug as an aid to psychotherapy. There were no rigorous clinical trials, but some users said that the drug helped them to trust others and to understand themselves better².

There are obvious parallels with the underground culture that surrounded LSD, or lysergic acid diethylamide, in the late 1960s and early 1970s. Although many users were simply intrigued by the hallucinations induced by the drug, some claimed that LSD sent them on a self-improving voyage of psychological discovery.

Doblin was part of this psychedelic underground. He grew up in a wealthy suburb north of Chicago, and started taking LSD in 1971 as a student at the New College of Florida in Sarasota, where students develop

individual programmes of study. At the time, Doblin was troubled by emotions that he attributes to the loss of some of his relatives in the Holocaust. "I had been suffering with the awareness that this scapegoating and cultural insanity was a direct threat to me," he says. "LSD just helped me put it all together." The experience also underpinned Doblin's self-directed major in 'transpersonal psychedelic psychotherapy'.

Going underground

Convinced of the healing power of psychedelic drugs, Doblin later began to train as an underground therapist, beginning by sitting in on therapy sessions aided by LSD

and MDMA. But his budding career as a therapist hit a snag in 1985, when a young graduate student named George Ricaurte reported that a related hallucinogen called MDA was toxic to rat neurons that produce serotonin, a neurotransmitter chemical that carries signals between brain cells that help control mood³. The US Drug Enforcement Administration promptly outlawed both MDA and MDMA.

This decision sparked Doblin to switch course to focus on legalizing MDMA and other psychedelic drugs. In 1988, he enrolled at Harvard University's Kennedy School of Government, where he earned a doctorate by studying the US medical regulatory system.



And he founded MAPS to fund research on the mental and physical effects of psychedelic drugs.

One of MAPS' earliest projects was a study in which MDMA was given to monkeys to examine its effects in the brain⁴⁻⁶. This project was led by Ricaurte, and the pair's fortunes have been tied together ever since. Initially, their relations were friendly — Doblin was even invited to Ricaurte's 1990 wedding to his colleague Una McCann. But more recently, they have become forthright opponents in the debate over the safety of MDMA, and whether it has a place in the psychiatric clinic.

Doblin believed that studies by Ricaurte and others would eventually demonstrate that MDMA was safe. But contrary to his hopes, this work — in particular results that have emerged from Ricaurte's lab at Johns Hopkins University in Baltimore, Maryland — has thrown up successive obstacles.

Most experts agree that MDMA decreases brain levels of serotonin in animals, and damages the branches through which serotonin-producing neurons reach out and connect to other cells⁷. It is difficult



Rick Doblin (left) and George Ricaurte (far right) have clashed over the safety of MDMA.



Pill power: ecstasy has been claimed by club culture as its drug of choice, but could it help to treat some psychiatric problems?

to determine whether the same changes happen in the brains of human MDMA users. But Ricaurte's group has used various indirect methods to find evidence that suggests they do, including measurements of a breakdown product of serotonin in cerebrospinal fluid⁸. Using a radioactive chemical that binds to serotonin receptors, the researchers have also shown from positron emission tomography brain scans that MDMA users have fewer receptors for the neurotransmitter⁹. "The bottom line is that over 15 years worth of research by us and others has clearly demonstrated that MDMA is toxic to brain serotonin neurons," McCann argues.

But the debate still rages over the long-term consequences of this damage. There are hints that MDMA affects processes such as sleep cycles¹⁰, but more alarming possibilities come from studies that indicate that MDMA users have poorer memory and impaired thought processes, compared with people who haven't tried the drug^{11,12}. The problem with ascribing these effects to ecstasy is that people who take MDMA also frequently stay up all night dancing and may also take other drugs.

The definitive test would be to give MDMA in a controlled setting to people who have never tried it before. In 1998, Franz Vollenweider of the Psychiatric University Hospital in Zurich, Switzerland, published a study in which he did just that, reporting on the drug's immediate cognitive effects, which included "moderate thought disorder"¹³. It was a highly controversial study, with critics including Ricaurte and McCann¹⁴ arguing that it was unethical to give people a drug that could cause permanent brain damage. Vollenweider did not study the treatment's long-term effects.

Amid this contentious debate, MAPS nevertheless managed to get the US Food and Drug Administration in 2001 to approve

a clinical trial of MDMA in patients with post-traumatic stress disorder. But in September 2002, Ricaurte, McCann and their colleagues published a paper in *Science*¹⁵ that seemed to spell the end for MDMA as a therapeutic tool. The researchers gave squirrel monkeys and baboons doses of the drug that they claimed were similar to those taken by some clubbers in a single night. This

killed neurons that produce the neurotransmitter dopamine, and led to a condition similar to Parkinson's disease. Two out of ten of the animals died.

"We already know from the literature that brain dopamine declines with age," Ricaurte said at the time. "A young individual who sustains injury to these

dopamine cells and depletes their reserve may be at greater risk of parkinsonism."

The resulting furor marked the nadir of Doblin's relationship with Ricaurte and McCann. After the paper had been submitted to *Science*, Doblin claims that McCann torpedoed MAPS' effort to start its clinical trial, which needed ethical clearance from an Institutional Review Board (IRB) before any patients were treated. The Western IRB, based in Olympia, Washington state, had told Doblin on 10 July 2002 that it would approve the study. But this approval was revoked on 4 September, says Doblin, after a doctor affiliated with the board spoke on the telephone with McCann.

Hard to swallow

Ricaurte's shocking findings were also bad news for Jose Carlos Bouso of the Autonomous University of Madrid in Spain. In late 2001, Bouso had earned regulatory approval and begun his own study of MDMA as a treatment for patients with post-traumatic stress disorder. Political pressure led Spanish drug-enforcement officials to halt the trial in May 2002, but Bouso was pressing to get it restarted. After

Can mushrooms work their magic?

While the controversy surrounding ecstasy has been grabbing the headlines, a quiet revolution in psychedelic-drug research has been under way. Over the past year, doctors have begun a handful of trials to test natural hallucinogenic drugs in psychiatric medicine.

Proponents of such trials say that Native American cultures have used such drugs for hundreds of years, and that doctors can build on this knowledge to learn how to use them. Psychiatrist John Halpern of Harvard Medical School in Boston has been studying the use of the hallucinogenic peyote cactus by the Native American Church, which holds that it is a religious sacrament handed down from God. The cactus contains the drug mescaline, and is taken during religious rituals to cure drug addiction and alcoholism. So far, says Halpern, preliminary observational studies have not shown the drug to have major side effects.

Meanwhile, Francisco Moreno of the University of Arizona at Tucson has treated eight patients with psilocybin, the active ingredient in 'magic' mushrooms (above right). He is testing anecdotal reports that the drug can help patients to manage symptoms of obsessive-compulsive disorder. And psychiatrist Charles Grob at the University of California, Los Angeles, is also testing psilocybin to relieve anxiety in terminally ill cancer patients.

Regulators have given these studies an easier time than those involving ecstasy



because mescaline and psilocybin don't seem to damage the brain, and also don't put people who take them under cardiovascular stress.

But even the proponents of such studies caution that the drugs are not risk-free. Moreno, for instance, says that psilocybin is safe when used under supervision in a controlled environment, but can cause intense and frightening hallucinations if people are not prepared for the experience. "If people aren't in the right mindset, hallucinations can be very threatening," he says. In his trial, Moreno is testing a range of doses to find out whether patients even need to experience psilocybin's full hallucinogenic effect to derive benefit from the drug.

publication of the *Science* paper, Doblin claims that Ricaurte impeded these efforts by making public speeches in Spain about the hazards of MDMA.

Ricaurte declined to be interviewed for this article, responding by e-mail: "I would think that it is preferable for a journal like *Nature* to focus on scientific issues related to MDMA, rather than personalities." But McCann argues that Doblin's complaints about her interaction with the Western IRB are based on a misunderstanding. She says that she did not discuss the results from the *Science* paper, and initially gave the trial her tentative backing, believing that it would be conducted in a hospital. But when she later learned that it would be run in a psychiatrist's office, she expressed misgivings. "I'd rather it be done in a hospital setting," McCann says. When the Western IRB withheld its approval, the trial's location was cited as one reason for the decision.

Doblin's crusade to bring MDMA into the clinic seemed to have finally been defeated. But in an embarrassing climb-down, Ricaurte and his colleagues were forced to retract their paper in September 2003, after discovering that the monkeys had actually been injected with methamphetamine, rather than MDMA (ref. 16). The

researchers blamed mislabelling by their suppliers at the National Institute on Drug Abuse in Bethesda, Maryland.

Just 18 days later, an IRB based in North Carolina stamped the MAPS study with its seal of approval. And on 24 February this year, the trial cleared its final hurdle when the Drug Enforcement Administration granted psychiatrist Michael Mithoefer a licence to dispense MDMA.

Each of the 20 patients in the trial will be given 11 hours of psychotherapy without any drugs. They will also undergo two different eight-hour sessions about a month apart, after which they will spend the night at Mithoefer's private clinic in Charleston. During these sessions, 12 patients will take MDMA, whereas eight will take a placebo.

A question of trust

Mithoefer, who also has an academic appointment at the Medical University of South Carolina, hopes that MDMA will help these patients to overcome their fears and trust him as their therapist. "They have trouble forming therapeutic relation-

ships," he explains. "Our hypothesis is that it may allow people to overcome those obstacles." But with so few patients enrolled in the study, this first trial won't produce a definitive answer. "This is a pilot study to see if there's any suggestion that MDMA can be useful," says Mithoefer. His first goal is to find out whether the drug is safe for people who have never taken MDMA before.

Apart from the question of whether the drug damages the brain, there are concerns about its cardiovascular effects. For instance, in 1996, Charles Grob, a psychiatrist at the University of California, Los Angeles, published a safety trial on 19 people, two of whom developed high blood pressure¹⁷. Grob had been considering giving MDMA to patients with end-stage cancer, to test whether it would help them to confront their anxiety about dying. But the political flap over ecstasy and the cardiovascular stress he observed has since made him turn instead to another psychedelic drug, psilocybin.

Doblin still has a lot of work to do if MDMA is ever to be embraced by the medical mainstream. But he sees the South Carolina trial as a key victory, and hopes that doctors and regulators will eventually change their minds if research proves that MDMA and other psychedelic drugs are safe and clinically effective. "We are at the forefront of studying risk and benefit," Doblin says. "We hope that there's enough respect for science that people will judge us on the basis of our data."

Erika Check is *Nature's* Washington biomedical correspondent.

1. Abraham, H. D., Aldridge, A. M. & Gogia, P. *Neuropsychopharmacology* **14**, 285–298 (1996).
2. Greer, G. & Tolbert, R. J. *Psychoactive Drugs* **18**, 319–327 (1986).
3. Ricaurte, G., Bryan, G., Strauss, L., Seiden, L. & Schuster, C. *Science* **229**, 986–988 (1985).
4. Ricaurte, G. A., DeLanney, L. E., Irwin, I. & Langston, J. W. *Brain Res.* **446**, 165–168 (1988).
5. Ricaurte, G. A. *et al.* *J. Am. Med. Assoc.* **260**, 51–55 (1988).
6. Ricaurte, G. A., DeLanney, L. E., Wiener, S. G., Irwin, I. & Langston, J. W. *Brain Res.* **474**, 359–363 (1988).
7. Green, A. R., Mehan, A. O., Elliott, J. M., O'Shea, E. & Colado, M. I. *Pharmacol. Rev.* **55**, 463–508 (2003).
8. McCann, U. D., Ridenour, A., Shaham, Y. & Ricaurte, G. A. *Neuropsychopharmacology* **10**, 129–138 (1994).
9. McCann, U. D., Szabo, Z., Scheffel, U., Dannals, R. F. & Ricaurte, G. A. *Lancet* **352**, 1433–1437 (1998).
10. Allen, R. P., McCann, U. D. & Ricaurte, G. A. *Sleep* **16**, 560–564 (1993).
11. Montoya, A. G., Sorrentino, R., Lukas, S. E. & Price, B. H. *Harv. Rev. Psychiatry* **10**, 212–220 (2002).
12. Morgan, M. J. *Psychopharmacology* **152**, 230–248 (2000).
13. Vollenweider, F. X., Gamma, A., Liechti, M. & Huber, T. *Neuropsychopharmacology* **19**, 241–251 (1998).
14. McCann, U. D. & Ricaurte, G. A. *Neuropsychopharmacology* **24**, 333–336 (2001).
15. Ricaurte, G. A., Yuan, J., Hatzidimitriou, G., Cord, B. J. & McCann, U. D. *Science* **297**, 2260–2263 (2002).
16. Ricaurte, G. A., Yuan, J., Hatzidimitriou, G., Cord, B. J. & McCann, U. D. *Science* **301**, 1479 (2003).
17. Grob, C. S., Poland, R. E., Chang, L. & Ernst, T. *Behav. Brain Res.* **73**, 103–107 (1996).

♦ www.maps.org

Biodiversity law has had some unintended effects

Moves to prevent unfair exploitation of resources could restrict conservation research.

Sir—Your News story “Bioprospectors hunt for fair share of profits” (*Nature* 427, 576; 2004) correctly draws attention to the bureaucratic impediments many countries have placed in the way of access to biodiversity in the wake of the Convention on Biological Diversity (CBD, 1992). It reveals that delays in developing benefit-sharing agreements are retarding the potential of biodiversity-rich countries to reap the full rewards of biotechnology. However, by overvaluing the commercial potential of biodiversity, CBD-based legislation in these countries is also impeding conservation science.

There was much hope after the CBD that developing countries would conserve biodiversity for its economic promise, specifically the potential pharmaceutical profits derived from biological resources. Such resources were called ‘green gold’, and these countries have been told, for example, that “people all over the world last year paid more than \$400 billion for pharmaceuticals, nearly half of which were discovered in the wild” (D. Labrador *Sci. Am.* 289, 17–18; December 2003). The media continue to publicize the handful of

glamorous examples of medicines obtained from tropical biodiversity — such as the appetite-suppressant properties of the Kalahari plant *Hoodia gordonii*.

Such hype has not been lost on the governments of developing countries, who have concluded that billions of ‘eco-dollars’ lie hidden in their forests. Understandably, they have responded by hastily framing laws to protect (rather than conserve) their biodiversity. Although they are mostly intended to facilitate access, many of these laws obstruct biodiversity-related research, rarely differentiating between commercial and conservation science. Meanwhile, commercial returns from benefit-sharing in these countries remain trivial compared with their national conservation budgets.

One of the first scientific victims of these restrictive regimes is taxonomy. Fewer than 10% of the species on Earth have been described, but access legislation in many developing countries alienates and criminalizes taxonomists, whose job it is to describe them. For example, regulations made under India’s Biological Diversity Act, passed in 2002, require any person

wishing to have “access to biological resources and associated knowledge for research” to seek government approval, paying an exorbitant US\$200 application fee (India’s annual gross national income per capita is \$460, according to the World Bank).

The adverse impact of the CBD on research is finally being more widely recognized. The International Union of Biological Sciences, at its 28th general assembly in Cairo in January 2004, having debated these issues at a one-day workshop, resolved to promote “activities that enhance scientific input to the CBD process”. These include the results of biodiversity inventories and population studies based on field exploration, which require access to the species concerned.

For many biologists working in developing countries, however, the jury is still out on whether the unintended negative consequences of the CBD outweigh its benefits.

Rohan Pethiyagoda

Adviser to the Ministry of Environment and Natural Resources, Sampathpaya, Battaramulla, Sri Lanka

Putting Norway on the gene-therapy map

Sir—I read with interest the Commentary by M. Cavazzana-Calvo and colleagues on “The future of gene therapy” (*Nature* 427, 779–781; 2004). However, readers should be aware that the statistics used to produce the map shown on page 781, “Number of approved gene-therapy trials”, are incomplete.

The map, using data compiled from available sources by the Wiley *Journal of Gene Medicine*, indicated that no gene-therapy trials have been performed in Norway. Yet, according to the Norwegian Directorate of Health, which approves such studies, seven trials were performed before 2004 — six at the Norwegian Radium Hospital and one at the National Hospital, both in Oslo.

For Norway, this is not just a question of incomplete statistics. When the first application for a gene-therapy trial was rejected by the Norwegian Board of Health in 1996, on the basis of sound scientific advice, rumours circulated in the international community that Norwegian authorities were hostile towards gene therapy. Your map supports this impression. Yet, during the past five years

the Norwegian Ministry of Health has funded an ambitious grant programme to help scientists in Norway acquire internationally competitive competence in gene therapy.

Some years ago, I participated in a technology-assessment study of gene therapy organized by the Norwegian Center for Health Technology Assessment, in which the Wiley, Medline, EMBASE and US National Institutes of Health databases were searched for information on clinical trials. This resulted in a much more complete survey, although there is still no single authoritative source for such information.

Ola Myklebost

Department of Tumour Biology,
Norwegian Radium Hospital,
N-0310 Oslo, Norway

Gene therapy needs both trials and new strategies

Sir—In their Commentary “The future of gene therapy” (*Nature* 427, 779–781; 2004), M. Cavazzana-Calvo and colleagues suggest that we proposed, in a recent Perspectives article (*Science* 302, 400–401; 2003), screening genetically modified stem cells before they are infused into patients,

to avoid potential serious side effects.

We are very much aware that such an approach is not applicable with current methods. Our comment represented our vision of where the field needs to move in the future. To thrive, the field needs both continuous development of realistic clinical trials — as outlined by Cavazzana-Calvo and colleagues — and conceptual discussions about new strategies. Regulatory decisions need to follow the former, scientific the latter.

As outlined in the Perspectives article, advanced cell culture, gene delivery and molecular screening technologies are likely to bring significant progress in the foreseeable future.

David Williams*, Christopher Baum†‡

*Cincinnati Children’s Hospital Medical Center,
Division of Experimental Hematology,
3333 Burnet Avenue, Cincinnati,
Ohio 45229-3039, USA

†Hannover Medical School, Department of
Hematology, Hemostaseology and Oncology,
Carl-Neuberg-Strasse 1,
30625 Hannover, Germany

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Making waves

The voyage of HMS *Challenger* launched the science of oceanography.

The Silent Landscape

by Richard Corfield

Joseph Henry/John Murray: 2003/2004.

304 pp. \$24.95/£20

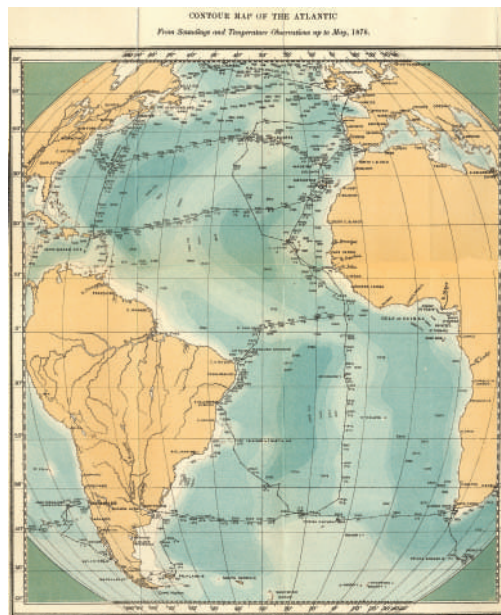
Charles Langmuir

In 1874 the *Challenger* expedition left England on the maiden voyage of global ocean science. The results still reverberate within the community of sea-going scientists: all research vessels have some version of the *Challenger* results in their small libraries, and the phrase “as first discovered by the *Challenger* expedition” is often heard. The voyage discovered life in the deep sea and the complex thermohaline structure of the ocean. It provided the first samples of rocks from the ocean basins. It made possible a global map of ocean depths. And it initiated virtually every field of oceanography, from rocks to water to life. The *Challenger* results dealt with all these issues and more, in 50 volumes published over a period of 25 years after the ship returned to England in 1878.

Less well known are the rigours associated with this famous voyage. Today, ocean-going research is well established and an international fleet of oceanographic vessels continually traverses the oceans. Some 20 scientists and an equal number of crew experience a month at sea with the benefits of perfect navigation, the engine power to go anywhere at any time, and daily communications with land.

The *Challenger* expedition was heroic by comparison. Close to 300 people, including only six scientists, were crammed on to a ship half the size of modern research vessels. The expedition was largely under sail, and for four long years the scientists and crew endured ferocious storms, demoralizing calms, threatening icebergs, plagues in foreign ports, the tragic deaths of colleagues and friends, and had only infrequent communication with home.

As well as the scientific and human stories, the expedition has a rich historical context. The voyage emerged from a confluence of practical, political and scientific motivations. The communications revolution of the telegraph led to the need for better maps of the sea floor, and part of the expedition was devoted to planning routes for telegraph cables. England's sea power provided the expertise, resources and support in foreign ports to mount such a major research effort. The debates about evolution ignited by Darwin's theory of natural selection raised the question of the missing links between species that might be found in the sea, and there was a hot debate about whether life in



In-depth study: Wyville Thomson's map shows that the *Challenger* expedition revealed the Mid-Atlantic Ridge.

the ocean was restricted to shallow depths where sunlight can penetrate. Scientific questions, practical needs, the projection of a global presence — all these could be related to subsequent global science initiatives.

Such a panorama opens a broad array of choices for the modern writer. Richard Corfield has elected to give us a short, readable volume that combines an abbreviated version of the expedition, snapshots of personal accounts of some of its participants, modern perspectives on selected scientific problems, and reflections on modern research vehicles in ocean drilling and space that have adopted the *Challenger*'s name.

The book is an entertaining and informative combination of history and modern science. It fits with the recent enthusiasm for books on historical geology, such as *The Map that Changed the World* (Viking, 2001) by Simon Winchester, and should appeal to the same audience. Many of the expedition's logistical details, showing the ingenuity and foresightedness of these pioneering oceanographers, will also amuse and appeal to the modern sea-going scientist.

The decision to keep the book short means that some scientific aspects of the expedition get little attention, and specialists may be disappointed by the superficiality of, and occasional scientific errors associated with, the treatment of some modern developments. This book nicely covers the modern outgrowth of climate science that stems from *Challenger*'s discoveries, but another author could equally have written a book

emphasizing the *Challenger* results that presaged plate tectonics or influenced our understanding of evolution and biodiversity. Corfield devotes only a few pages to plate tectonics, and refers disparagingly to the “tedium of dredging and sounding”. For those who appreciate it, however, scientific dredging is almost always exciting, as it penetrates the unknown depths and brings back from the bottom materials never before seen by human eyes. It was dredging that brought back the sediments, provided the first samples of the basalts that make up the ocean ridges, and recovered the thousands of new species documented by *Challenger*'s biologists.

Can we learn anything today as we reflect on the discoveries of the *Challenger* expedition? Perhaps they serve to remind us of the limits of our imagination, and that the answers we receive are often limited by the questions we are able to ask. In the late

nineteenth century, Earth was thought to be static, with limited horizontal motion. Continental drift had not yet been proposed, so questions of movement and of changes to the shape of the ocean basins never came to mind. And yet Wyville Thomson, who died of exhaustion shortly after the expedition ended, may have had some premonition of what was to come. An elegant frontispiece in his book summarizing the expedition reveals a map of the Mid-Atlantic Ridge, fully recognizable in remarkable detail. Thomson also noted without further comment that the ridge closely matched the shape of the coastlines on the two sides of the Atlantic. The facts were there; the questions were not. Similar observations, and the more detailed map produced by Bruce Heezen and Marie Tharp some 80 years later, were central to the discovery of plate tectonics, which revealed that the solid Earth is continually in movement and circulation. Which of today's observations will prove most interesting when we know what questions to ask?

One characteristic of a successful book is that one puts it down longing for more. Within its chosen limits, *The Silent Landscape* is indeed successful. It provides the history and excitement of an epic voyage in the context of modern developments. It does so in a brief and readable form, and leaves ample scope for deeper explorations of such rich historical material. ■

Charles Langmuir is in the Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

Exhibition

A window on the past

Glassware used to be very precious but, as with all commodities, more efficient production methods made the price tumble. This is what happened in Roman times.

Inventories from houses in Pompeii show that by 79 AD — when volcanic ash from Vesuvius smothered the city — the middle classes had replaced most of their metal household vessels with glass ones. Glass was still too expensive for poor households, but was apparently too cheap for the rich, who seemed to have stuck with their gold and silver vessels.

The exhibition "Vitrum", which runs until 31 October in Florence, shows how Roman technologists made glass affordable — in large part by perfecting glass-blowing techniques, which vastly increased its versatility too. They also developed other techniques, such as glass 'cameoing', and colouring using different metal oxides.

With its new and seductive transparency, glass quickly won over many levels of society. In



the Pompeii mosaic above, the fruit bowl, whose fruit is sparkingly visible within it, is clearly a source of domestic pride. But glass also transformed architecture. Many glass window panes have been found in Pompeii, replacing the traditional cloth hangings that left interiors dark. Mosaics for walls then became popular, using coloured glass chips which had reflective

properties that further lightened domestic space.

The scientific élite also embraced glass for its own pragmatic and experimental purposes. Physicians began to use glass urine containers. Urine was a major diagnostic aid in those days, and a transparent vessel that allowed its colour to be easily monitored was a boon. Sealed glass ampoules also came into common use to protect pharmaceutical preparations from contamination.

In the second century AD, Ptolemy used variously shaped vessels — "made of glass that is

as thin and pure as possible so they are transparent" — that he filled with water, to demonstrate his laws of optical refraction.

This unusual and informative exhibition enjoys an exceptional location in Florence's fifteenth-century Palazzo Pitti, in rooms whose walls are completely frescoed. **Alison Abbott**
<http://brunelleschi.imss.fi.it/vitrum>

Exponent of the exponential

The Man Who Shocked the World: The Life and Legacy of Stanley Milgram

by Thomas Blass

Basic Books: 2004. 368 pp. \$26, £19.99

Steve Blinkhorn

It's a sin to tell a lie — or is it? Experimental psychology has always engaged in modest amounts of subterfuge to distract human subjects from the true purpose of investigations. Experimental social psychology, or at least one branch of it, goes much further and makes deceit in social interaction the pith and essence of investigative practice. The usual social norms of sincerity, spontaneity and good faith are discarded to provoke attitudes or behaviour in unsuspecting human subjects: this is true experimental manipulation in the interests of a science of social behaviour.

Stanley Milgram was an outstanding exponent of this style of research, but to what end? His famous experiments on obedience, where subjects were duped into believing they were administering painful and even dangerous electric shocks to other volunteers in the course of a banal learning experiment, both launched and blighted his professional career. Like so much else of his

work, they made manifest a sublime talent for dramatizing and presenting what was already known about the moral frailty of humankind. Who since Shakespeare, or even Sophocles, has believed that human behaviour is governed by the rational evaluation of alternatives under the guidance of a well-examined conscience? Lawyers, perhaps. Milgram found ways of capturing this frailty without adding substantially to our understanding of it, or to our ability to remedy it.

The writer of this welcome but flawed biography, Thomas Blass, thinks otherwise. The author of more papers on Milgram than Milgram himself ever published, he veers from pointed critical evaluation of Milgram's work and character to overblown hagiographic praise. We are invited to think of Milgram as great, with an ever-present sense of muffled outrage at the failure of Harvard University to grant him tenure. The puffery extends to a dust-jacket endorsement by a professor of physics, who extols Milgram as the most influential social psychologist of the past century. When did physicists ever look to psychologists to evaluate the standing and contribution of their own peers?

A careful reading reveals that Milgram was a chippy colleague, often unpleasant to his students, and had two characteristics often taken as warning signs of unreliability: a drug habit and a sense of entitlement.

In fact, for all his undoubted talent for communication to a wider public, Milgram

singularly failed to win research grants from major funding bodies, saw remarkably few students through to completion of doctoral studies, and turned increasingly to the media as an outlet for his talents. He lived his research methods. When he burst into a seminar with the news of John F. Kennedy's assassination, it was simply assumed that this was Milgram up to his usual tricks, trying another experimental social manipulation.

The other major contribution attributed to Milgram, beyond his obedience work and some technical methods for unobtrusive research, concerns the idea of six degrees of separation: that anyone can find a chain of acquaintance with anyone else on Earth that has about six links. This idea has ramifications not just in the study of social networks, but also in computing and mathematics. It is little elaborated in this work, perhaps because Milgram himself did little to elaborate it, beyond the bold statement published in the first issue of *Psychology Today*.

But how bold was it? To untutored common sense it may seem surprising, but untutored common sense falls prey to chain letters and pyramid selling schemes, which share essentially the same mathematics, and has little grasp of the reach of exponential arithmetic. To make six degrees of separation a fair representation of reality, on a hopelessly simple model (and the interesting issues are all to do with why such a simple model is inadequate) at the time Milgram was writing, the average number of unique

acquaintance links per person was 42. This perhaps casts the Meaning of Life, the Universe and Everything in a new light. Our talent for procreation, however, has subsequently increased the number to 43.

This is an uneven book — perhaps an overzealous editorial hand has been at work. You will feel at home reading it if you know the detailed topography of Harvard University, what a summer salary is and why it is important, the precise nature of a proseminar, the US system of academic appointment and tenure, and East Coast Jewish culture; otherwise you may just feel a slight sense of social exclusion. On the other hand, perhaps unintentionally, the book offers an insight into a world whose public outings are usually more carefully groomed and polished in the interests of impression management.

And of course, this being *Nature*, all of the above statements are true. Or are they? ■

Steve Blinkhorn is at Psychometric Research and Development, Brewmaster House, The Maltings, St Albans, Herts AL1 3HT, UK.

Studying form

Jacob's Ladder: The History of the Human Genome

by Henry Gee

Fourth Estate: 2004. 272 pp. £20

To be published by Norton in the US in July

Andrew Berry

Jacob's Ladder opens with an account of human embryonic development. It's a wonderful piece of writing, its lyricism inspired perhaps by Henry Gee's choice of his own daughter as the focal fetus. And it nicely sets up the book's theme: the "question of what it is that produces form from the formless". What follows is an engaging mix of history of science, evolutionary biology and molecular genetics. Much of it is familiar: we see Mendel pottering in the monastery garden, Darwin being seasick on the *Beagle*, and *Hox* gene products binding to DNA regulatory regions. But Gee, a senior editor at *Nature*, has stitched the material together in interesting ways and has plenty of points of his own to make.

One of Gee's messages is that, although "the story of biology can be told in an unbroken skein", the discipline's intellectual history is wilfully ignored. Too often the publication of Darwin's *The Origin of Species* in 1859 is deemed the starting point of modern biology, to the extent that "if we hear anything at all of biology before Darwin, it is brought up only to be belittled." Gee, then, is on a mission to restore the reputations of the pre-darwinians, and the first part of the book duly pushes a 'nothing new under the sun' thesis.

Take preformationism, the doctrine that



Unravelling the spiral: after sequencing the human genome, how will we attempt to manipulate it?

gametes contain preformed miniature humans (homunculi) and that development is simply a matter of amplifying what is already there. It all seems wonderfully silly to us, especially as the logical corollary is that Eve's ovaries contained, ready formed, the germ of every human to come. But Gee believes that modern biology's proudest achievement, the Human Genome Project, owes much to preformationism. After all, under both preformationism and modern notions of development, "conception does not start life from scratch, but simply activates a program that was already in existence" — that's the genome if you're in the twenty-first century, or the preformed germ if you're in the eighteenth.

Gee is keen that we recognize our debt to our intellectual antecedents, but his purpose is not to induce a spasm of ancestor worship; rather, he sees history as rich in lessons. The genome project was peddled as latterday preformationism — know the sequence, or homunculus, and all of humanity's secrets will be revealed — but the shortcomings of such a conveniently one-dimensional perspective are today amply apparent as pages of journals become clogged with talk of transcriptomes and proteomes, and nobody, to my knowledge, has yet fingered the consciousness gene.

That humanity's secrets have proved less accessible than the hype suggested, Gee believes, could have been predicted by paying proper attention to previous scientific controversies. His respect for the past is laudable, but I need more convincing that intellectual history should be a necessary part of a scientist's education. The example given here is more a condemnation of the rhetoric associated with the genome project than of the scientific work itself. The press and science writers certainly oversold the project, and many scientists participated in the hype. But this wasn't science, it was PR. For

such an expensive and high-profile undertaking, PR, like it or not, is a necessary evil.

Jacob's Ladder does not confine itself to the past; indeed the ladder of the title, which in the original in Genesis was populated by angels en route to heaven, symbolizes for Gee a future in which we manipulate our own genome. Whether the resulting ladder will ultimately lead up or down is surely the question, but Gee does not dwell on this. His vision for the future is more short term and more securely scientific: his strongest message is his endorsement of a network approach to biology.

The key to understanding how living systems work lies in uncovering patterns of interaction among genes and their products. Gee is an enthusiastic proponent of what these days is billed as 'systems biology': a combination of high-tech data gathering, computer simulation and big-picture thinking. The ability to study many genes at once and the availability of computers that can perform once-inconceivable combinatorial feats do indeed together represent an exciting opportunity. But, to return to Gee's "unbroken skein" of the history of biology, I don't see it as a major departure from what has gone before.

Population geneticists — a species that seems to try Gee's patience — have long grappled with multilocus problems. With biochemists, they have long studied the properties of a particular kind of network, the biochemical pathway. Even the currently fashionable developmental networks were extensively treated by population geneticists such as C. H. Waddington more than 50 years ago. Gee is absolutely correct to point to network analysis as a major way forwards. But it is ironic, in view of his embrace of biology's past, that he does so with a lack of regard for what came before. ■

Andrew Berry is at the MCZ Labs, Harvard University, Cambridge, Massachusetts 02138, USA.

Heads and tails

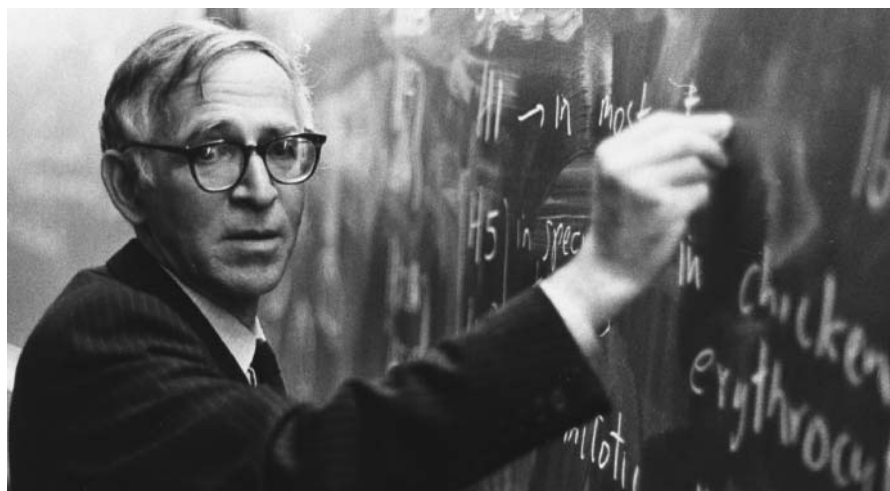
How a visit to the home of structural biology inspired a young scientist.

Mitsuhiro Yanagida

I first visited Cambridge in June 1968 to attend the International Symposium on Interactions Between Subunits of Biological Macromolecules. It was an excellent meeting, with talks by prominent researchers — I was particularly pleased to hear Francis Crick, who often led the discussions. A mere eight months before this meeting, I had joined Eduard Kellenberger's group at the Institute of Molecular Biology in the University of Geneva, Switzerland. Eduard saved me from being a graduate student in University of Tokyo — the riots in Japanese universities erupted soon after I joined him, and many of my generation were unable to continue their studies.

My work involved using phage morphogenesis as a model for more complex biological structures. My first project was to use an optical diffractometer to analyse the images of electron micrographs showing the tubular 'polyheads' produced in bacteria that were infected with phage T4 mutants. The polyhead was an ideal structure to study with the diffraction analyses of electron micrographs, a technique developed by Aaron Klug in the Laboratory of Molecular Biology (LMB), Cambridge. Although electron microscopy had a theoretical resolution at the atomic level, actual micrographs of biological specimens rarely achieved this because of deformation, damage, noise and the superimposition of images at different focal planes. However, if the biological particles consisted of regularly repeated structures, diffraction analysis could be a powerful tool for obtaining information from the average lattice structure. Another technique, optical filtration, also enabled selection for the side image of helical structures. I inherited the polyhead project from another student, who had left to become an activist for the independence of Jura. In retrospect, social unrest was responsible for two lucky events in my early scientific career.

I obtained a large number of diffraction patterns from polyhead electron micrographs. It seemed clear to me that the polyhead was not the precursor of the normal T4 head shell, as its lattice structure was polymorphic; the polyhead was therefore probably an abnormal assembly. I gave a short talk on these findings at a specialist workshop organized by Aaron after the main symposium. He told me that David DeRosier had found the same things in his laboratory, but that my micrographs might be good enough for further analysis by the optical-filtration method. Although Aaron had already pioneered the



The enthusiasm of Nobel prizewinner Aaron Klug added to the unique atmosphere at Cambridge.

computational Fourier reconstruction for the three-dimensional phage tail structure from electron micrographs, he was still interested in studying the head-related structures by manual diffraction methods.

Aaron and Eduard decided to submit an application for a short-term European Molecular Biology Organization (EMBO) fellowship, which would enable me to work at the LMB for a few months, and I arrived there on a cold, rainy day in December 1968. The laboratory had been established in 1962 by the UK Medical Research Council, following the pioneering work in Cambridge of the previous decade — which included the discovery of the structure of DNA by Watson and Crick. Many of the staff (Max Perutz, Crick and Fred Sanger) were already Nobel laureates and others would soon follow.

On the first day, I was nervous and afraid that I might not understand what Aaron would talk about. From his famous papers on the X-ray crystal analysis of tobacco mosaic virus (initially working with Rosalind Franklin), the quasi-equivalent bonding in the icosahedral virus shell (with Don Caspar) and helical diffraction theory (with Crick), I was convinced that he was a theoretical physicist. I was concerned that with my poor knowledge of diffraction theory, I might be asked to return to Geneva immediately.

To my great surprise, Aaron was interested in experimental details and interpretations, and spent most of the time talking about how to get a clue or a starting point for determining the assembly mechanism of T4 head proteins. I was relieved to find myself discussing ideas with him. He was an extremely good physics teacher, and a fantastic experimentalist. He had the amazing ability to analyse details of various images by eye, and

was able to uncover so many things from raw data. Aaron also spoke of his broad interests in science — beside those of the principal projects of his group. It is no wonder that he was later successful in elucidating the structures of transfer RNA, the nucleosome and the zinc-finger transcription factor, among others. I did not realize at the time, but he had already completed the studies for which he would get a Nobel Prize. Through Aaron's ability and interest, I understood for the first time the depth and strength of structural molecular biology at the LMB. My stay was over all too soon, but by then we had produced several very beautiful filtered images of polyheads.

By merely breathing the air of the LMB, something ultimately changed in my mind and flesh. Everywhere, the spirit of freedom in scientific thinking prevailed — in the seminar room, cafeteria and even the corridors. People there simply exchanged scientific ideas on equal terms. The LMB looked to me like a paradise for the elite scientists who worked there. A second short visit to Cambridge later strengthened this feeling.

I began to have an increasing affinity for British ideas on philosophy, science and life. In my short time working with Aaron at the LMB, I was inspired to make a career as a scientist. However, it took six more years and my own tiny research group in Kyoto University to determine how the three distinct subunits were arranged to form the normal T4 head shell. After that peak of excitement, and a brief intermission with the visualization of single DNA molecules in solution, I encountered the fission-yeast chromosome — which was to be my playground for next 25 years. ■

Mitsuhiro Yanagida is at the Graduate School of Biostudies, Kyoto University, Yoshida-Honmachi, Kyoto 606-8501, Japan.

A fuel-efficient geodynamo?

Richard Holme

The hunt has been on for a source of extra power to account for the dynamo that generates the Earth's magnetic field. A synthesis of computation and experiment now suggests that the search may not be necessary after all.

The Earth's magnetic field originates in the core, a fluid region almost 3,000 km below our feet of which the main constituent is molten iron. The basic mechanism of field generation is generally accepted — it is a dynamo process, in which the magnetic field is maintained by convection of the highly electrically conducting fluid. But the details are less clear.

For example, how much power is required to drive the geodynamo? With current estimates, the inferred heat flow from the core is not enough to drive a dynamo over most of Earth's history, as is required from measurements of a strong magnetic field recorded in ancient rocks. One possible explanation lies in a re-evaluation of the composition of the core — could it contain more potassium, which would provide an additional heat source from radioactive decay¹? Some geochemists are coming round to this idea, but it remains controversial. On page 169 of this issue², Christensen and Tilgner offer an alternative — they argue that less power than previously estimated is required to drive the geodynamo, and that no additional heat sources are necessary.

The past ten years have seen great progress in understanding the generation of the geomagnetic field, driven primarily by improvements in computing resources that have made numerical models of the dynamo possible. Some dynamo models closely mimic features seen in magnetic-field observations, both historical and over millions of years^{3,4}. However, because computational resources are inadequate, and are likely to remain so, none of these models can claim to represent the Earth: model viscosity values are orders of magnitude too high for the core, and the rotation speeds are much too slow. As a result, direct quantitative comparison with the Earth is impossible. In fact, we are so far from using true Earth-like parameters that the surprise is not that the models don't always produce fields that look like the Earth's, but that any of them do.

To address this problem, two complementary approaches have emerged. The first is to drive the numerical models as close to Earth-like parameters as possible, often with 'unphysical' numerical schemes, and to hope that the resulting models are close enough to realistic conditions. The second has been to consider a suite of dynamos in more computationally accessible regimes and, by

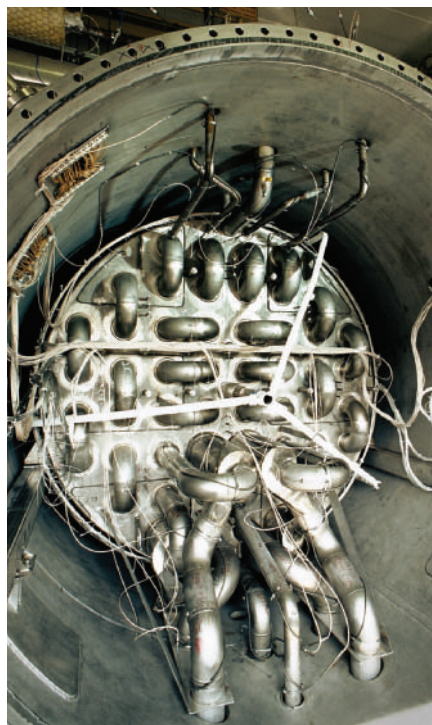


Figure 1 Core simulations: the innards of the Karlsruhe experimental dynamo. Liquid sodium is driven through the 10-cm-diameter pipes, and the flow generates and sustains a magnetic field.

varying the parameters involved (viscosity, rotation rate, convection strength), to derive scaling relationships that can then be extrapolated to an Earth-like regime. However, the relationships are derived over little more than one order of magnitude of parameter space, and must then be extrapolated over many more orders of magnitude to reach that regime.

Christensen and Tilgner² demonstrate precisely the pitfalls of such a scheme. They obtain two scaling relations for the power required to drive the dynamo, one a function only of the strength of the flow, and a second additionally a function of fluid viscosity. Both models fit their simulations well; the second model is only weakly dependent on viscosity, but for extrapolation to realistic Earth conditions that dependence is vital. Which model should we believe: the one-parameter model, which allows the geodynamo to be driven without an additional heat source, or the two-parameter model, which requires more power?

Christensen and Tilgner address this

issue by recourse to a new and exciting tool for studying the core. Experimental dynamos are being developed by several groups⁵, involving apparatus that contains highly conducting liquids, usually liquid sodium. One such experiment⁶ in Karlsruhe, Germany, has successfully generated a magnetic field by dynamo processes in a parameter regime inaccessible to numerical models (Fig. 1). The results from that experiment are consistent with the simpler, one-parameter scaling relation, but not with the additional viscosity dependence. On this basis, the authors argue for a lower energy requirement to drive the dynamo, so that there is no need for extra potassium in the core.

It would be easy to poke holes in this analysis. The conclusion depends crucially on the experimental results. Although the experimental dynamo has a viscosity appropriate for the core, in other ways it is far from Earth-like. For example, rather than being a homogeneous sphere of fluid, it uses a complicated system of pipes to produce a flow geometry suitable for dynamo action. And it is kinematic: that is, the fluid is pumped, and the feedback of the magnetic field on the flow is very limited.

It could therefore be argued that the results have nothing to do with the geodynamo. However, other dynamo experiments are under development, consisting of contained spheres of liquid sodium, that are much closer to Earth's core geometry⁷. As results become available from these experiments, Christensen and Tilgner's approach will become increasingly powerful. We will probably never be able to model the parameter regime of the Earth, either by computation or experiment. But perhaps by combining the two we may be able to 'sneak up' on the Earth and make true quantitative predictions for the geodynamo. ■

Richard Holme is in the Department of Earth and Ocean Sciences, University of Liverpool, 4 Brownlow Street, Liverpool L69 3GP, UK.
e-mail: holme@liv.ac.uk

1. Nimmo, F., Price, G. D., Brodholt, J. & Gubbins, D. *Geophys. J. Int.* **156**, 363–376 (2004).
2. Christensen, U. R. & Tilgner, A. *Nature* **429**, 169–171 (2004).
3. Bloxham, J. *Geophys. Res. Lett.* **29**, 1854 (2002).
4. Glatzmaier, G. A. & Roberts, P. H. *Nature* **377**, 203–209 (1995).
5. Gaillitis, A., Lielausis, O., Platacis, E., Gerbeth, G. & Stefani, F. *Rev. Mod. Phys.* **74**, 973–990 (2002).
6. Müller, U. & Stieglitz, R. *Nonlin. Proc. Geophys.* **9**, 165–170 (2002).
7. Lathrop, D. P., Shew, W. L. & Sisan, D. R. *Plasma Phys. Contr. F* **43**, A151–A160 (2001).

Signal transduction

Thumbs up for inactivation

Holger Rehmann and Johannes L. Bos

Cellular signalling pathways depend on the proper activation and inactivation of mediators. New structural information reveals an unusual mechanism by which one such mediator, Rap1, is switched off.

Proteins of the Ras family regulate a variety of cellular processes — ranging from proliferation to adhesion — in response to information received at the cell surface. These proteins, which are part of a broader class known as G proteins, are in essence simple on/off switches. Whether they are on or off is determined by the small molecules to which they bind: they are on when bound to guanine nucleotide triphosphate (GTP), and off when bound to guanine nucleotide diphosphate (GDP). The switch is flicked on by guanine-nucleotide-exchange factors (GEFs), which exchange bound GDP for GTP, and off by GTPase-activating proteins (GAPs), which accelerate the ability of Ras proteins to hydrolyse GTP into GDP (Fig. 1a). On page 197 of this issue, Wittinghofer and colleagues¹ present the crystal structure of one of these GAPs, revealing a new and unusual mechanism by which Ras-family proteins can be inactivated. This is the latest in a series of investigations by the same group that aims to unravel how GAPs work^{2,3}.

The importance to cells of being able to inactivate Ras proteins is immediately evident from the fact that cancer is associated with mutations that render the founding member of the family, Ras itself, permanently active: such mutations are found in 30% of all metastatic tumours⁴. These mutations occur at amino-acid positions 12, 13 and 61 of the protein, and prevent the GAP for Ras (RasGAP) from doing its job.

Wittinghofer and colleagues previously carried out a detailed structural analysis² of the complex between Ras and its GAP, trapped in the process of GTP hydrolysis (the 'transition state'). The structure revealed the molecular mechanism by which the GAP induces GTP hydrolysis, and why cancer-causing mutations in Ras render the protein incapable of being inactivated. Simply put, RasGAP works by inserting an arginine amino acid into the nucleotide-binding pocket of Ras, completing the catalytic machinery (Fig. 2a). This amino acid has been dubbed the 'arginine finger', and it has two functions. First, it neutralizes the negative charge that develops at the γ -phosphate — the phosphate group in GTP that will be removed during hydrolysis into GDP. Second, the arginine finger fixes the conformation of the glutamine at position 61 in Ras. Glutamine 61, in turn, positions a water molecule ready to attack the γ -phosphate.

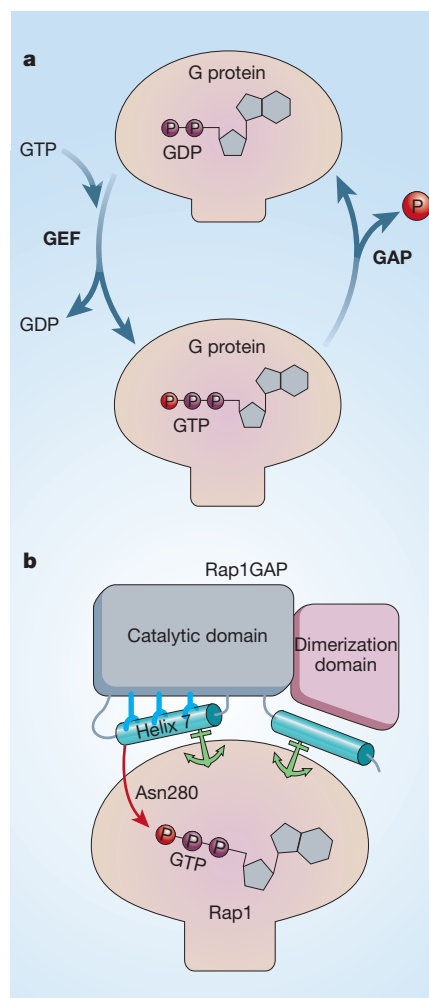


Figure 1 Switching G proteins on and off. a, G proteins, such as members of the Ras family, cycle between a GDP-bound, inactive state and a GTP-bound, active state. Transition from the active to the inactive state is achieved when the G protein hydrolyses its GTP. This intrinsically slow process is accelerated by GTPase-activating proteins (GAPs). Transition to the active state requires exchange of GDP for GTP, a process catalysed by guanine-nucleotide-exchange factors (GEFs). b, The complex of Rap1 and its GAP, showing amino acids in the GAP that Wittinghofer and colleagues¹ find to be important for its activity. Blue, amino acids required to maintain the protein's structural integrity; green anchors, amino acids involved in binding to Rap1; red arrow, asparagine (Asn) 280 — an amino acid that participates directly in the catalytic process, shown attacking the phosphate in GTP to be removed (the γ -phosphate).

This, then, explains the cancer-promoting nature of mutations at position 61: the arginine finger doesn't have the same effect on amino acids other than glutamine, and those other amino acids can't position the water molecule correctly. As to the effect of mutations at positions 12 and 13, any amino acid other than glycine at these positions interferes with the hydrolysis reaction by displacing the arginine finger and thereby glutamine 61.

The interplay between an arginine finger and a glutamine has been assumed to be at the heart of all GTP-hydrolysis reactions. Indeed, the active centre of the Ras-family protein Rho and its GAP is set up in exactly the same way⁵, although RhoGAP and RasGAP are structurally unrelated. And an arginine and glutamine with the same function have been found for G α proteins, the only difference here being that the arginine is part of the G protein itself⁶. The complex of Ran and RanGAP³ shows some differences, but is essentially a variation on the same theme: RanGAP does not provide an arginine, but it still fixes the conformation of glutamine 61 of Ran (Fig. 2b).

The closest relative of Ras is Rap1, which regulates the adhesion of cells to one another or to surrounding matrices⁷. This protein is the odd one out in the Ras family, in that it lacks a glutamine at position 61. So the stimulation of GTP hydrolysis in Rap1 must involve a substantially different mechanism. To get a handle on this mechanism, Wittinghofer and colleagues¹ have now solved the crystal structure of Rap1's GAP. They find that it is structurally unrelated to any other known GAP. The protein consists of a catalytic domain and a dimerization domain (Fig. 1b); the latter domain allows Rap1GAP dimers to form and also stabilizes the catalytic domain.

To identify the amino acids involved in catalysis, the authors carried out a mutational analysis of the protein, looking at the effects of altering specific amino acids. They found that mutations that abolish or weaken the affinity of Rap1GAP for Rap1 were clustered in and around helix 7 of the GAP. The amino acids here either stabilize helix 7, as suggested by the crystal structure, or are exposed to the solvent, and so, the authors suggest, might form the surface that interacts with Rap1.

The consequences of mutating asparagine 280 were particularly striking. This mutation did not weaken the affinity of Rap1GAP for Rap, but totally abolished its catalytic activity. The function of asparagine 280 as the catalytic amino acid is proven by the inability of transition-state analogues to stabilize a complex between the mutated Rap1GAP and Rap1. (Transition-state analogues mimic the chemical nature of GTP during hydrolysis, and bind to Ras-like proteins only when all catalytic amino acids are present.) In analogy with the arginine finger, Wittinghofer

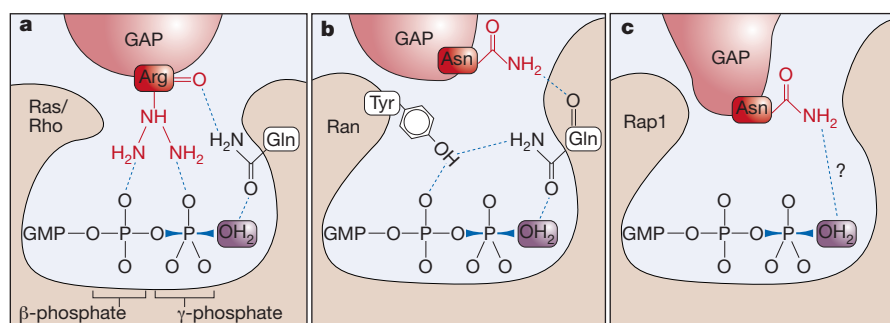


Figure 2 The active sites of some G-protein-GAP complexes. The complexes are shown in the transition state through which they pass during GTP hydrolysis. A common task for all GAPs is to position and fix a water molecule (purple) that attacks the γ -phosphate of GTP. **a**, Ras in complex with RasGAP², or Rho in complex with RhoGAP⁵. Here, the 'arginine finger' (Arg) in the GAP helps to stabilize the glutamine (Gln) in the G protein, which in turn positions and fixes the H₂O. The arginine finger also neutralizes the negative charge that develops at the γ -phosphate, by contacting this and the β -phosphate. **b**, Ran in complex with RanGAP³. RanGAP does not provide an arginine finger, instead using an asparagine (Asn) to stabilize the glutamine. The negative charge is neutralized by a tyrosine (Tyr) in Ran itself. **c**, Rap1 in complex with Rap1GAP. Wittinghofer and colleagues¹ propose that here, an asparagine (Asn) in the GAP positions and fixes the H₂O directly. GMP, guanosine monophosphate.

and colleagues¹ call asparagine 280 the 'asparagine thumb', and propose a model in which the attacking water molecule is positioned directly by this thumb (Fig. 2c). We look forward to seeing this model confirmed by a structure, mimicking the transition state, of Rap1 in complex with its GAP.

There are other proteins related to Rap1GAP, notably Tsc2. This protein is mutated in tuberous sclerosis — a benign type of human tumour — and functions as a GAP for the Ras-family member Rheb⁸. Whereas Rap is unique in lacking glutamine 61, Rheb is unique because the glycine at position 12 is replaced by an arginine. This is incompatible with the mechanism adopted by Ras and its GAP, but perhaps the asparagine thumb is the solution in this case, too. Indeed, an asparagine corresponding to that at position 280 of Rap1GAP is mutated in people with tuberous sclerosis⁹.

It remains a mystery why Rap1 uses a different mechanism of GAP-induced GTP hydrolysis from other members of the

family. Perhaps it is a way to maintain fidelity between G proteins — which are generally extremely similar — and their respective inactivators. Indeed, the introduction of a glutamine at position 61 makes Rap sensitive to another GAP entirely — that for Ras¹⁰.

Holger Rehmann and Johannes L. Bos are in the Department of Physiological Chemistry and the Centre for Biomedical Genetics, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands. e-mail: j.l.bos@med.uu.nl

1. Daumke, O., Weyand, M., Chakrabarti, P. P., Vetter, I. R. & Wittinghofer, A. *Nature* **429**, 197–200 (2004).
2. Scheffzek, K. *et al. Science* **277**, 333–338 (1997).
3. Seewald, M. J., Korner, C., Wittinghofer, A. & Vetter, I. R. *Nature* **415**, 662–666 (2002).
4. Bos, J. L. *Cancer Res.* **49**, 4682–4689 (1989).
5. Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gambin, S. J. *Nature* **389**, 758–762 (1997).
6. Coleman, D. E. *et al. Science* **265**, 1405–1412 (1994).
7. Bos, J. L., de Rooij, J. & Reedquist, K. A. *Nature Rev. Mol. Cell Biol.* **2**, 369–377 (2001).
8. Zhang, Y. *et al. Nature Cell Biol.* **5**, 578–581 (2003).
9. Maheshwar, M. M. *et al. Hum. Mol. Genet.* **6**, 1991–1996 (1997).
10. Hart, P. A. & Marshall, C. J. *Oncogene* **5**, 1099–1101 (1990).

Quantum physics

High NOON for photons

Dirk Bouwmeester

Entangled photons conspire to create interference patterns that would normally be associated with a wavelength much smaller than that of the individual photons — beating the diffraction limit.

If a double slit is illuminated with a laser beam of wavelength λ , the familiar rippling pattern of interference fringes arises, even if the attenuation of the laser is so strong that only single photons pass through the double slit at any given instant. To emphasize this quantum feature, Paul Dirac wrote¹ that "Each photon interferes only with itself.

Interference between two different photons never occurs." Dirac would probably have been more cautious had he read this issue of *Nature*. Starting on page 158, Mitchell *et al.*² and Walther *et al.*³ report their demonstrations of interference patterns produced by specific entangled states of three and four photons.

The resolution in each case is better than it would be for a single photon passing the interferometer: $\lambda/3$ and $\lambda/4$ in the three- and four-photon case, respectively, where λ is the wavelength of an individual photon. Normally, resolution no better than the photon wavelength can be achieved, which is known as the diffraction limit. So, as well as illustrating the properties of multi-particle entanglement, these limit-breaking results are of practical interest — for high-resolution optical read-out and recording systems, for example, and interferometric position and time measurements.

To illustrate how entanglement can be used to produce an interference pattern with a resolution of λ/N (where N is the number of entangled particles), consider first the simplest case of $N=2$. Two photons are entangled such that they go together through slit A or through slit B of a double-slit configuration. A shorthand notation for this state is $|2_A 0_B\rangle + |0_A 2_B\rangle$ or, even more compact, $|2, 0, 2\rangle$. That this state indeed represents an entangled state follows from the fact that it cannot be written as a product state of two individual photons going through the double slit. The product state would have the additional possibility of one photon going through slit A and the other through slit B. In essence, the two entangled photons form one system, a 'bi-photon' if you like, that has twice the energy of a single photon and, therefore, half its wavelength. If a photographic plate that is only sensitive to bi-photons is positioned behind the double slit, an interference pattern associated with the wavelength $\lambda/2$ would be observed.

States of the form $|2, 0, 2\rangle$, with corresponding $\lambda/2$ resolution of the interference pattern, have been demonstrated using photons produced by the optical process of 'parametric down-conversion'^{4–6}. Here, a photon from a pump laser is converted into two photons, each with half the energy of the initial photon. Although quantum physicists might get quite excited about such proof-of-principle experiments, the large community of scientists working on high-resolution measurement systems might feel somewhat cheated. After all, those experiments start with a pump laser with wavelength $\lambda/2$ to produce the fancy entangled photons at wavelength λ that then lead to a $\lambda/2$ interference pattern. Why not just use the pump-laser photons directly and obtain the same resolution?

It would be more interesting if $|N, 0, N\rangle$ states could be generated with $N>2$ but using photons produced by light sources that have a wavelength of at least $\lambda/2$. The existence of such states — dubbed 'high NOON' states by Jonathan Dowling — would be an unambiguous demonstration that the diffraction limit has been beaten. This is exactly what Mitchell *et al.*² and Walther *et al.*³ have achieved, with $|N, 0, N\rangle$ states for $N=3$ and $N=4$, respectively.

Mitchell *et al.* have shown that a $|30,03\rangle$ state can be created by probabilistically superposing three photons with specific polarizations. The experiment closely follows the proposal by Hofmann⁷ and is related to several other proposals^{8–10}. An interference pattern with resolution $\lambda/3$ is obtained by detecting three-photon coincidences. Walther *et al.* present a variation of the $|40,04\rangle$ state. They show how double pairs of polarization-entangled photons emerging from two separate (but phase-related) sources can be combined to give a pure four-photon interference pattern. A clever choice for the detection of four-photon coincidences reveals the desired interference pattern with a resolution of $\lambda/4$.

Both schemes^{2,3} have proved in an elegant way that the diffraction limit can be beaten, and both are in principle scalable to even higher resolutions. However, from a practical point of view, the signal-to-noise ratio of the interference fringes should also be analysed. Currently, it requires very long measurement times to resolve three- and, in particular, four-photon interference fringes with rather low signal-to-noise ratios. Single-photon interference fringes can quickly be detected with much higher signal-to-noise ratios,

resulting in higher accuracy for, for example, a position measurement. To push the 'high NOON' technology forward, it will be crucial to develop bright entangled multi-photon sources and efficient multi-photon detectors. Looking some years ahead, it doesn't seem unrealistic to expect commercial applications of entangled photons in secure data transmission systems, high-resolution optical read-out and storage systems, and perhaps even quantum computation. As if the world isn't entangled enough! ■

Dirk Bouwmeester is in the Department of Physics, University of California, Santa Barbara, California 93106, USA.

e-mail: bouwmeester@physics.ucsb.edu

1. Dirac, P. A. M. *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1982).
2. Mitchell, M. W., Lundeen, J. S. & Steinberg, A. M. *Nature* **429**, 161–164 (2004).
3. Walther, P. *et al.* *Nature* **429**, 158–161 (2004).
4. Ou, Z. Y., Zou, X. Y., Wang, L. J. & Mandel, L. *Phys. Rev. A* **42**, 2957–2965 (1990).
5. Rarity, J. G. *Phys. Rev. Lett.* **65**, 1348–1351 (1990).
6. Fonseca, E. J. S., Monken, C. H. & Padua, S. *Phys. Rev. Lett.* **82**, 2868–2871 (1999).
7. Hofmann, H. F. Preprint at <http://arxiv.org/abs/quant-ph/0311198> (2003).
8. Pryde, G. J. & White, A. G. Preprint at <http://arxiv.org/abs/quant-ph/0304135> (2003).
9. Kok, P., Lee, H. & Dowling, J. P. *Phys. Rev. A* **65**, 052104 (2002).
10. Fiurasek, J. *Phys. Rev. A* **65**, 053818 (2002).

Membrane trafficking

Dual-key strategy

Toshiki Itoh and Pietro De Camilli

Traffic flow between cellular compartments is controlled by recruitment of cytoplasmic proteins. New work exemplifies a dual-key mechanism, involving membrane lipids and proteins, that coordinates this control.

Our cells contain a series of distinct compartments that do different jobs and have different properties. The membranes that clad each of these compartments — like the plasma membrane that encases the cell — are defined by precise molecular compositions, which are preserved despite the continuous influx and efflux of components in transit to and from other cellular locations. Precision is the hallmark of this flow of traffic, too, which must be directed appropriately between compartments. All of this is achieved, in part, by the reversible recruitment of regulatory proteins from other parts of the cell to specific membranes or membrane regions. A growing amount of evidence hints that membrane lipids cooperate with membrane proteins to control this recruitment. In *Nature Cell Biology*, Godi and colleagues¹ reveal an example of this cooperative spirit, which helps to ensure the proper progression of cargo from one compartment — the *trans*-Golgi network (TGN) — to the cell surface.

Phosphatidylinositol is one of the most common lipids in cellular membranes, and

generally localizes to the cytosolic side of a membrane (the side that faces into the cell). The reversible addition of phosphate groups to (that is, phosphorylation of) the inositol ring of this lipid, at the 3, 4 or 5 positions, generates seven phosphoinositide species in mammalian cells. Each species recognizes, with varying affinities and specificities, certain modules or amino-acid motifs that are found in cytosolic proteins. So, in a mechanism reminiscent of the phosphorylation of tyrosine amino acids in membrane proteins², differential phosphorylation of the inositol ring helps to control the recruitment of specific proteins to membranes. In turn, the distribution of different phosphoinositides to different membranes depends on the distribution of the enzymes that add or remove phosphate groups^{3,4}.

Two of the proteins to which phosphoinositides bind are FAPP1 and FAPP2 (FAPP stands for 'four-phosphate-adaptor protein'). These are members of a protein family that is thought to function in the sorting or metabolism of lipids⁵. Both proteins have a so-called pleckstrin homology (PH) domain,



100 YEARS AGO

The death of Sir H. M. Stanley on Tuesday, at sixty-three years of age, deprives the world of a man of action, and geography of one of its greatest pioneer explorers. It can truly be said that he changed the map of Africa by the results of his expeditions, and his picturesque narratives created public interest in the problems of African exploration. Stanley's adventures in Central Africa while engaged in the search for Livingstone attracted great attention, and his famous book, "How I Found Livingstone," in which the expedition is described, has become a classic work of travel. Commissioned to find Livingstone, of whom nothing had been heard for two years, Stanley reached Zanzibar in January, 1871, and on November 10 of the same year met the explorer at Ujiji, on Lake Tanganyika, where Livingstone had just arrived from Nyangwe... In 1874, Stanley left England for the expedition to Central Africa which has immortalised him. The writer of the obituary notice in the *Times*, from which some of the particulars here given have been derived, points out that little more than the position of Victoria Nyanza was then known; ... our knowledge of Albert Nyanza was incomplete; Lake Tanganyika was imperfectly defined; and nothing was known of the region that lies between Lakes Albert and Tanganyika. Stanley's expedition changed all that. From *Nature* 12 May 1904.

50 YEARS AGO

While, however, from the point of view of advancing knowledge, the hydrogen bomb and like experiments are necessary and are being carried on with due regard to all reasonable safeguards, both legal and scientific, it may still be asked whether we have not in fact reached, or approximated to, the point of no return. The purpose of continued experiment is to prevent any potential enemy of the free world from gaining a completely decisive lead. It may be argued that atomic weapons have already reached a stage at which it is meaningless to speak of a decisive lead. Here it needs to be remembered again that the limited resources of Britain do not permit us to provide our Armed Forces with all those things which they would wish to have... The tests are extremely costly and laborious, and those responsible for carrying them out are acutely aware of the pressure for financial economies, which is just as strong in the United States as in Britain. From *Nature* 15 May 1954.

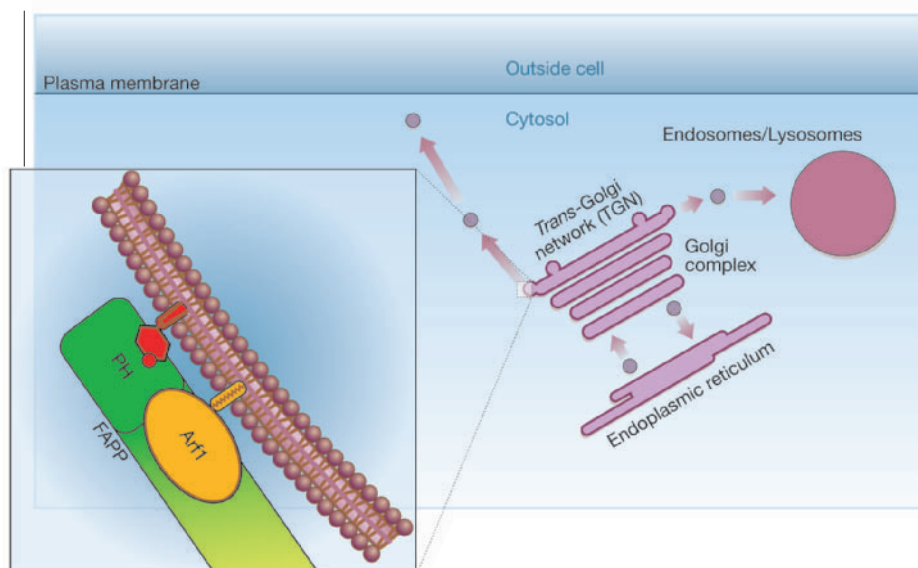


Figure 1 Mechanism for the recruitment of the proteins FAPP1 and FAPP2 to the *trans*-Golgi network (TGN). Some of the main compartments within cells are shown, namely the endoplasmic reticulum, the Golgi complex (the part of the complex that is farthest from the endoplasmic reticulum being the TGN) and endosomes/lysosomes. Vesicles carry cargo from one compartment to another. Godi *et al.*¹ propose that the FAPPs participate in vesicular transport from the TGN to the cell surface. The binding of these proteins to TGN membranes is mediated by their pleckstrin homology (PH) domains (see inset), through the interaction of this domain with the lipid phosphatidylinositol-4-phosphate (red) and the protein Arf1. If one of these binding partners is missing, FAPPs no longer localize to the TGN.

which in this case binds specifically to phosphatidylinositol-4-phosphate (PtdIns(4)P) — a phosphoinositide involved in the formation of membrane-clad organelles, such as vesicles and tubules, that carry cargo from the TGN^{3,6–8}.

Godi and co-workers¹ now show that both FAPP1 and FAPP2 are found at the TGN (Fig. 1). More notably, when the expression of these proteins is prevented, the transport of cargo from this compartment to the cell surface is impaired. Conversely, when cells are forced to produce too much of the isolated PH domain of these proteins, long, narrow tubules, which do not break off into vesicles, emerge from the TGN. These tubules contain cargo that is destined for the cell surface, and extend towards the plasma membrane but do not fuse with it. The authors interpret the appearance of these tubules as resulting from competition between the overexpressed PH domains and the cell's own PtdIns(4)P-binding proteins — possibly the FAPPs themselves — that participate in the creation and breaking-off of vesicles destined for the cell surface.

Godi *et al.* also show that the localization of the FAPPs to the Golgi complex is prevented by a mutation in the PH domain that stops it from binding to PtdIns(4)P, as well as by inhibiting the production of this phospholipid. Moreover, in agreement with previous studies⁵, the authors find that the PH domain alone can localize to the TGN, indicating that this lipid-binding domain contains all the information needed for such localization.

How exactly the FAPPs assist in the formation of carriers destined for the cell surface remains uncertain, although the presence of a glycolipid-transfer domain in FAPP2 points to a potential link to lipid metabolism^{1,5}. Regardless, Godi and colleagues' findings are significant because, until now, molecules that bind PtdIns(4)P and explain its previously discovered role of directing traffic from the Golgi complex to the cell surface^{3,6,7} had not been identified. Another protein implicated in vesicular traffic at the TGN, clathrin adaptor-1 (AP-1), binds PtdIns(4)P (ref. 8) but does not participate in trafficking to the cell surface.

The new work also explains the very selective targeting of the FAPPs (and their PH domains) to the Golgi complex, despite the presence of PtdIns(4)P in other membranes. Previous studies^{5,9} had shown that the PH domain of oxysterol-binding protein — a member of the same protein family as the FAPPs — binds to PtdIns(4)P and localizes mainly to the Golgi complex. But a mutation in this particular PH domain that prevented it from binding to the lipid did not stop it from being targeted to the TGN. In yeast, this lipid-independent localization required the function of Arf1, a protein of the small-GTPase family that localizes to Golgi membranes⁵. It remained unclear, however, whether this effect of Arf1 was direct or indirect.

Godi *et al.* show that the PH domains of the FAPPs bind directly to Arf1 as well as to PtdIns(4)P. Moreover, although a single

mutant PH domain that cannot bind to PtdIns(4)P no longer localizes to the Golgi complex, this localization is restored when the mutant PH domain is expressed as a dimer, because the two Arf1-binding sites from the two PH domains cooperate. Arf1 also interacts with, and recruits to the Golgi complex, an enzyme that phosphorylates the 4 position of phosphatidylinositol¹⁰ — so Arf1 can lead to FAPP recruitment through at least two independent, but synergistic, mechanisms. Interestingly, the binding of AP-1 protein to membrane also requires Arf1 (ref. 11), hinting that PtdIns(4)P and Arf1 cooperate in several membrane-trafficking events at the TGN.

The discovery of protein modules that bind selectively to specific phosphoinositides has prompted the use of fluorescently tagged modules as 'reporters' of the distribution of the cognate phospholipids in living cells¹². These probes have greatly advanced our understanding of phosphoinositide signalling. However, as Godi and colleagues' study shows¹, these modules might contain binding sites for membrane proteins as well, so one cannot rely on such probes to reveal the entire distribution of a given phosphoinositide. Dual interactions with phosphoinositides and other membrane determinants have been detected for other PH domains too^{1,13}. Furthermore, a recent survey of all of the PH domains in budding yeast revealed that the subcellular targeting of these domains was only partly correlated with their phosphoinositide-binding specificities *in vitro*¹⁴. And similar dual interactions might apply to other phosphoinositide-binding modules as well, not just PH domains¹⁵.

The interaction of cytosolic proteins with both lipids and proteins on a target membrane is an efficient dual-key strategy to control their recruitment to membranes. Only when both the lipid-binding and protein-binding sites are engaged is the interaction with the membrane strong enough. The two elements of the code can be controlled independently, affording the possibility of fine-tuning the spatial and temporal regulation of recruitment. Small GTPases and phosphoinositides have emerged as important switches in the regulation of membrane interfaces. This latest study¹ provides yet another example of how the two machineries cooperate with each other.

Toshiki Itoh and Pietro De Camilli are at the Howard Hughes Medical Institute and Department of Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06510, USA.
e-mail: pietro.decamilli@yale.edu

- Godi, A. *et al.* *Nature Cell Biol.* **6**, 393–404 (2004).
- Schlessinger, J. *Cell* **103**, 211–225 (2000).
- Odorizzi, G., Babst, M. & Emr, S. D. *Trends Biochem. Sci.* **25**, 229–235 (2000).
- Wenk, M. R. & De Camilli, P. *Proc. Natl Acad. Sci. USA* (in the press).

5. Levine, T. P. & Munro, S. *Curr. Biol.* **12**, 695–704 (2002).
6. Walch-Solimena, C. & Novick, P. *Nature Cell Biol.* **1**, 523–525 (1999).
7. Hama, H. *et al.* *J. Biol. Chem.* **274**, 34294–34300 (1999).
8. Wang, Y. J. *et al.* *Cell* **114**, 299–310 (2003).
9. Levine, T. P. & Munro, S. *Curr. Biol.* **8**, 729–739 (1998).
10. Godi, A. *et al.* *Nature Cell Biol.* **1**, 280–287 (1999).
11. Stamnes, M. A. & Rothman, J. E. *Cell* **73**, 999–1005 (1993).
12. Balla, T. & Varnai, P. *Science STKE* doi:10.1126/stke.2002.125.pl3 (2002).
13. Varnai, P. *et al.* *J. Biol. Chem.* **277**, 27412–27422 (2002).
14. Yu, J. W. *et al.* *Mol. Cell* **13**, 677–688 (2004).
15. Chidambaram, S., Mullers, N., Wiederhold, K., Haucke, V. & von Mollard, G. F. *J. Biol. Chem.* **279**, 4175–4179 (2004).

Physiology

Orphan detectors of metabolism

Steven C. Hebert

There are myriad G-protein-coupled receptor proteins in living organisms, but the functions of many are unknown. Two of them are now shown to provide a link between metabolism and blood pressure.

One of the first things that biology students are taught is how cells liberate usable energy from food and oxygen. In eukaryotic cells (loosely speaking, those that have nuclei), much of this process of respiration occurs in double-membrane-bounded compartments known as mitochondria. A crucial stage in respiration is the tricarboxylic acid (TCA) cycle or Krebs cycle¹, into which are fed the products of the breakdown of sugars, fats and proteins, and out of which comes reducing power, used to generate the energy-storing molecule ATP. As testimony to the importance of this process, mutations in human genes encoding TCA-cycle enzymes have been associated with inherited cancers and severe generalized organ abnormalities². It has been known for some time that TCA-cycle intermediates are also present outside cells, in the blood circulation, and that extracellular levels of these organic acids increase when tissues are damaged by low oxygen supplies or other 'insults'. On page 188 of this issue, He and colleagues³ show, unexpectedly, that two such intermediates act as signalling molecules, linking the metabolism and injury of tissues with blood pressure.

Cells can detect and respond to a wide range of external signals by using receptor proteins on their surfaces. G-protein-coupled receptors (GPCRs) are the largest and most diverse group of such detectors; as their name suggests, they transmit extracellular information into cells by coupling to G proteins — common molecular switches — in the cytoplasm⁴. Although all GPCRs have similar structures, they show considerable diversity at the sequence level. However, on the basis of specific sequence motifs, they have been divided into three distinct families, each containing subfamilies that often have related stimuli (ligands)^{5,6}.

The number of GPCRs is growing continually: many new putative receptor genes have been identified in genome databases by computer-based screening for similar sequences. But for many of the newly

discovered receptors, a physiologically relevant ligand is unknown^{6–8}. These 'orphan' receptors are of considerable interest to biologists because they might define new ways in which cells can respond to their external environment, and to the pharmaceutical industry because they might provide new targets for drug development. Finding the relevant ligand for an orphan receptor can be a daunting task, however. One approach has been to use phylogenetic analyses — based on similarities in amino-acid sequence — to place newly identified receptors into one of the GPCR subfamilies, in an attempt to identify the ligand by the company the receptor keeps. But this is frequently not successful, and finding the correct ligand often requires a wide-ranging screening approach.

This was just the case for the orphan receptors GPR91 and GPR99. A phylogenetic approach had hinted that these proteins might bind nucleotides, given their membership in the P2 purinoceptor subfamily. He *et al.*³ have now found, however, that nucleotides do not activate these receptors. Instead, the TCA-cycle intermediates α -ketoglutarate and succinate are physiologically relevant ligands for GPR99 and GPR91 respectively.

What might these metabolic detectors do? Kidneys express both GPR91 and GPR99 (refs 7, 8), and He *et al.* provide compelling evidence that, via GPR91, succinate stimulates the release of renin from the kidneys. This reveals the mechanism behind the earlier observation that succinate increases the release of renin from isolated kidney capillaries⁹. Renin is part of the renin-angiotensin system (RAS) — a series of enzymes and molecules that culminate in the production of the eight-amino-acid peptide angiotensin II. This molecule in turn acts on blood vessels, causing them to constrict and so raising blood pressure.

Physiologically speaking, this could all be very useful. Mitochondrial dysfunction, caused by, for instance, an imbalance between energy demand and the food and oxygen

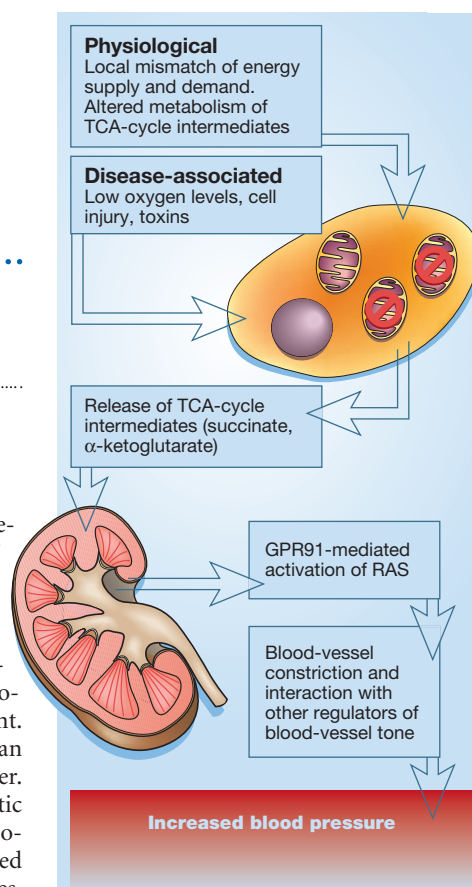


Figure 1 Connecting metabolism to blood supply. He *et al.*³ have found that the 'orphan' G-protein-coupled receptors GPR91 and GPR99 detect succinate and α -ketoglutarate — intermediates produced by the tricarboxylic acid (TCA) cycle during respiration. They also find that succinate increases blood pressure in mice. The figure shows how this happens. A local mismatch of energy supply and demand, altered metabolism of TCA-cycle intermediates, or injury leads to mitochondrial dysfunction and the release of succinate and α -ketoglutarate. These molecules activate receptors in the kidney, causing the release of renin and activation of the renin-angiotensin system (RAS). The RAS leads to an increase in blood pressure and altered local blood flow. Physiologically, this system might act to regulate local blood flow to match metabolic demands. However, it might also result in hypertension or alter cellular function.

supply, increases the extracellular concentrations of TCA-cycle intermediates (Fig. 1). He and colleagues' findings show that these intermediates can then influence the RAS, possibly (among other effects) reducing blood flow in the kidneys. Kidneys receive more than their fair share of the cardiac output of blood (about 20%), so reductions in kidney blood flow provide a means of redirecting blood to other organs during crises.

On the other hand, the balance can also be tipped towards disease: He *et al.* found that treating mice with succinate led to high

Muscle

The sliding filament at 50

"In spite of the numerous investigations which have been made into the changes of the striations of muscle when it contracts, there is little agreement at the present day on either the nature or the significance of these changes." Thus started the first of two independent, ground-breaking papers^{1,2}, published together in *Nature* on 22 May 1954, which brought general agreement about those changes. The papers showed that muscle shortens by relative sliding between two sets of subcellular filaments containing the proteins myosin and actin. This was the first demonstration that a cell's primary function could be understood in terms of a fundamental interaction between two protein molecules.

In the first of the papers¹, Andrew Huxley and Rolf Niedergerke reported measurements of the optically dense 'A-bands' in intact fibres from striated (skeletal) muscle. Using a novel interference microscope, the authors demonstrated that the width of the A-bands remains constant during contraction. To account for their observations, they suggested a 'sliding-filament' model



in which myosin filaments run the length of the A-band and actin filaments slide into this band when muscle shortens.

In the second paper², Hugh Huxley (no relation to Andrew) and Jean Hanson described their light-microscope investigations of isolated myofibrils from striated

muscle; myofibrils are subfibres of muscle that are thinner and more suitable for light microscopy. Huxley and Hanson independently established the constancy of the A-band width and also invoked a sliding-filament model to explain their data. They also extracted myosin from the

A-bands and demonstrated the role of ATP hydrolysis in powering the contraction-relaxation cycle of muscle.

Pictured here are the four protagonists. Clockwise from bottom left: Andrew Huxley, Rolf Niedergerke, Hugh Huxley and Jean Hanson.

The impact of the early work, and later developments in understanding the molecular mechanisms of what have since become known as motor proteins, are the subject of two meetings^{3,4} to be held in London next week. The two classic papers are reproduced in full in a special web focus⁵. The web focus also includes selected *Nature* publications that subsequently advanced our understanding of the molecular basis of muscle contraction and its bearing on an intriguing issue — the biological conversion of chemical energy into mechanical work. **Maxine Clarke**

1. Huxley, A. F. & Niedergerke, R. *Nature* **173**, 971–973 (1954).
2. Huxley, H. E. & Hanson, J. *Nature* **173**, 973–976 (1954).
3. www.royalsoc.ac.uk/events/discussion_meetings/level_2/may04.html
4. www.kcl.ac.uk/175/events/event0520.html
5. www.nature.com/nature/focus/slidingfilaments

blood pressure (hypertension) through the activation of the RAS. A similar mechanism could perhaps account for, or contribute significantly to, the renin-mediated hypertension that is associated with constriction of the renal artery by cholesterol plaques, as occurs in diabetes, kidney disease and other conditions. If the TCA receptors are indeed involved, however, then receptor blockers might be expected to inhibit kidney renin secretion and relieve hypertension.

And there are further possible medical implications. The concentrations of TCA-cycle intermediates depend to a certain extent on the state of the enzymes that are responsible for their turnover. Succinate levels, for instance, might be increased when the mitochondrial activity of the complex containing succinate dehydrogenase and its activator, coenzyme Q₁₀ (CoQ₁₀), is low. Low concentrations of CoQ₁₀ occur after heart failure and hypertension. Oral administration of this coenzyme improves heart function in some patients¹⁰, and some patients with hypertension can discontinue anti-hypertensive medications after taking CoQ₁₀ (ref. 11). It would be interesting to know

whether these effects are linked to a decrease in extracellular concentrations of succinate and hence in the activation of the RAS. Another observation is that some drugs use succinate as a salt; could this be having an effect in some patients?

TCA-cycle intermediates such as α -ketoglutarate and succinate have also been shown to reduce the mitochondrial injury, in kidneys and other tissues, that is caused by successive decreases and increases in oxygen concentrations¹²; such intermediates are also used in kidney (and liver) preservation solutions for organ transplantation¹³. What role, if any, might the TCA receptors have here? And might these receptors function in tissues other than kidneys? The complete RAS is expressed in many tissues, including the placenta, brain, heart, gonads and pancreas^{14,15}. Intriguingly, the placenta also expresses GPR99 (ref. 7), and disturbances in placental RAS activity can reduce placental blood flow in pregnancy, often with severe complications¹⁴.

Many aspects of the function of the newly identified TCA receptors clearly remain to be determined. Similar exciting

discoveries are likely to come as more and more orphan receptors find their homes in the GPCR family. ■

Steven C. Hebert is in the Department of Cellular and Molecular Physiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520-8026, USA.
e-mail: steven.hebert@yale.edu

1. Barnett, J. A. *Yeast* **20**, 1015–1044 (2003).
2. Eng, C., Kiuru, M., Fernandez, M. J. & Aaltonen, L. A. *Nature Rev. Cancer* **3**, 193–202 (2003).
3. He, W. *et al.* *Nature* **429**, 188–193 (2004).
4. Horn, F. *et al.* *Nucleic Acids Res.* **31**, 294–297 (2003).
5. Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. *Nature Rev. Mol. Cell Biol.* **3**, 639–650 (2002).
6. Joost, P. & Methner, A. *Genome Biol.* **3**, RESEARCH0063 (2002).
7. Wittenberger, T. *et al.* *BioMed Central Genomics* **3**, 17 (2002).
8. Wittenberger, T., Schaller, H. C. & Hellebrand, S. *J. Mol. Biol.* **307**, 799–813 (2001).
9. Baumbach, L., Leyssac, P. P. & Skinner, S. L. *J. Physiol. (Lond.)* **258**, 243–256 (1976).
10. Jeejeebhoy, F. *et al.* *Am. Heart J.* **143**, 1092–1100 (2002).
11. Hodgson, J. M., Watts, G. F., Playford, D. A., Burke, V. & Croft, K. D. *Eur. J. Clin. Nutr.* **56**, 1137–1142 (2002).
12. Feldkamp, T. *et al.* *Am. J. Physiol. Renal Physiol.* **286**, F749–F759 (2004).
13. St Peter, S. D., Imber, C. J. & Friend, P. J. *Lancet* **359**, 604–613 (2002).
14. Nielsen, A. H., Schauser, K. H. & Poulsen, K. *Placenta* **21**, 468–477 (2000).
15. Lavoie, J. L. & Sigmund, C. D. *Endocrinology* **144**, 2179–2183 (2003).

Biophysics

How DNA avoids getting wound up

Preprint at <http://arxiv.org/abs/q-bio.SC/0404037> (2004)

There is only one way to untangle DNA from its tightly packed *in vivo* form, ready for transcription. Maria Barbi and colleagues reach this conclusion by studying the topological constraints on plant- and animal-cell chromatin fibres, in which DNA is wound around protein discs to form a string of bead-like nucleosomes. This is just one of the levels of hierarchical structure in chromatin (the DNA-protein composite of our genetic material), but its topological characteristics determine the way DNA becomes unpacked so that the genetic data are readable.

Chromatin fibres are typically organized into long loops pinned at each end, and provided that the fibre is not snipped apart by enzymes, unpacking part of a loop must preserve a topological quantity called the linking number, which depends on how the fibre is twisted around itself like a bundle of string. Barbi *et al.* show that there is just a single unwinding pathway that takes a fibre from a compacted form as dense as that seen in the cell to an open form, while conserving the linking number. Chromatin seems to have evolved precisely the nucleosome spacing that allows this unfolding pathway to be accessed. **Philip Ball**

Cardiovascular disease

The genetics of risk

Hum. Mol. Genet. **13**, 993–1004 (2004)

Our risk of developing cardiovascular disease is thought to depend partly on our levels of 'good' and 'bad' cholesterol (that is, on levels of high- and low-density lipoproteins, HDLs and LDLs, respectively). These concentrations in turn depend on both genetic and environmental factors. Hans Knoblauch *et al.* have found that most of the genetic differences in people's cholesterol levels can be explained by changes in only 13 genes that are involved in lipid metabolism.

The authors measured the cholesterol levels of more than 1,000 people in 250 German families. They correlated these measurements with 93 variations, called single nucleotide polymorphisms (SNPs), in the 13 genes; these SNPs could be divided into groups, or haplotypes, that tend to be inherited together. Knoblauch *et al.* found that around one-third of the variation in cholesterol levels can be attributed to genes, and the rest to environmental factors. Variation in the haplotypes identified explained 67% of the genetic variation in LDL levels, 58% of that in HDLs and 99% of that in the HDL/LDL ratio.

Should the finding hold true in other populations, the next step will be to identify specific haplotypes that are linked to unhealthy cholesterol levels and that might be used to identify people at risk. **Helen Pearson**

Biomechanics

In a flap

J. Fluid Mech. **506**, 147–155 (2004)

Large flying animals, as well as some that swim, do so by flapping their forelimbs up and down. Microbes, in contrast, use tiny oars or spinning molecular motors to propel themselves along. Why the difference? The traditional answer is that larger creatures have to fight much harder against gravity. But how, then, do they create forward thrust while staying airborne?

Nicolas Vandenbergh and colleagues addressed the question by studying a model wing as it was moved up and down at varying frequencies in a tank of water. At low frequencies, the wing simply flapped vertically. But above a certain threshold it also began to move horizontally, and unidirectionally, during each stroke.

This process creates asymmetrical patterns of fluid movement on either side of the wing, giving it distinct 'leading' and 'trailing' edges. The authors acknowledge that animals will deliberately angle their wings or flippers. They nonetheless argue that, with increasing flapping frequency, there is a natural transition towards the development of thrust and the onset of horizontal motion, which occurs with even the simplest of wing motions. **Michael Hopkin**

Geochemistry

Ancient air not a fire hazard

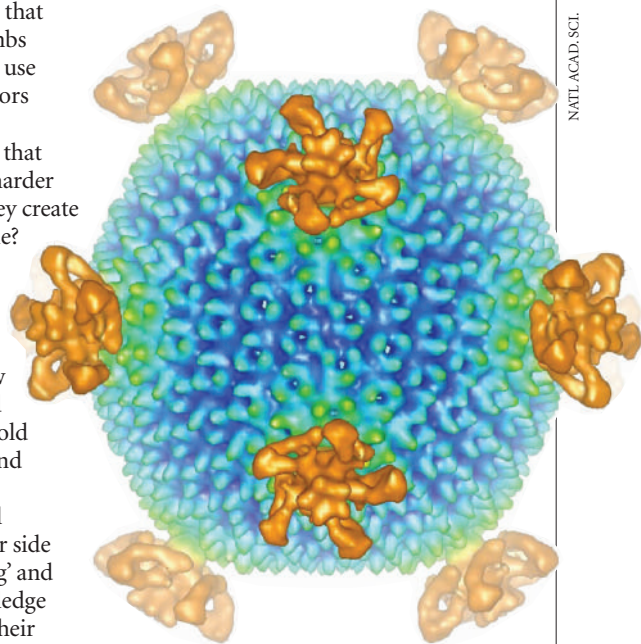
Geology **32**, 457–460 (2004)

Prehistoric forests may have been less combustible than has been thought, according to experiments by Richard A. Wildman Jr and colleagues. These findings challenge the notion that wildfires caused by high levels of atmospheric oxygen could have been catastrophic.

It has been argued that biological feedbacks tightly constrain both the upper and lower limits for the atmosphere's oxygen content: too little, and aerobic respiration becomes untenable; too much (more than 25% or so), and natural wildfires, ignited by lightning, could devastate plant life. That idea sits uncomfortably with geochemical models that suggest the atmosphere contained up to 35% oxygen about 300 million years ago.

To resolve the issue, Wildman *et al.* looked at the effect of differing ambient oxygen levels on the burning of wood, foliage and moss chosen to approximate the vegetation of Carboniferous forests.

Under realistic fuel moisture conditions, the combustion rates and fire-spreading propensity do not seem to be sufficiently increased by oxygen contents of up to 35% for wildfires to have reached disastrous proportions. Previous experiments used paper, which burns more readily than moist vegetation, and so gave misleading results. **Philip Ball**



Virology

High-temperature infection

Proc. Natl Acad. Sci. USA

doi:10.1073/pnas.0401773101 (2004)

There are three domains of life — the Archaea, Bacteria and Eukarya — and all three can be infected by specific viruses. Comparatively little is known about the viruses that infect the Archaea, a group of tiny microbes. But George Rice *et al.* have now analysed the genome and protein structure of a heat-loving archaeal virus (see picture) that lurks in the hot springs of Yellowstone National Park.

The authors isolated the virus from its host, *Sulfolobus solfataricus*, then extracted and sequenced its DNA and analysed its protein make-up. They found that the viral genes are quite different from those of other viruses and other organisms — perhaps indicating that this virus has a unique lifestyle compared with other species, or that it has been ecologically isolated for a lengthy period. But despite the genetic diversity, the structure of the viral 'coat protein' resembles that of bacterial and eukaryotic viruses.

The finding hints that some viruses may have a common ancestor that predated the split into the three domains of life more than 3 billion years ago. Analysis of this and similar viruses should, the authors say, allow new insights into the Archaea. **Helen R. Pilcher**

NATL ACAD. SCI.

Ageing and the mystery at Arles

Shino Nemoto and Toren Finkel

What determines how long we will live? Studies of simple organisms, single cells and mammals hint that certain shared principles underlie ageing, and raise the possibility of devising ways to extend life — if we want to.

"I don't want to achieve immortality through my work. I want to achieve it through not dying" — Woody Allen (1935–?)

In August 1997, Jeanne Calment was laid to rest in a simple ceremony in the south of France. Those who eulogized her life that day noted that it was unremarkable in all but one respect: she had been born on 21 February 1875. Five years before her birth, Napoleon III had been overthrown and the French Third Republic was formed. In her native Arles, Impressionism would take root and flourish. Years later, she would often tell visitors of her recollection of the painter Vincent Van Gogh, whom she regarded as "dirty and badly dressed". When pressed for the secret of her longevity, Madame Calment would stress her consumption of large quantities of olive oil and port wine, and the fact that she quit smoking, albeit aged 119.

Insurance companies study life expectancy all the time (see Box 1), but science long neglected the issue. Although the events that initiate life and the processes that determine its end were avidly examined, there was curiously little study of what determines the speed at which we travel between these two points.

Ageing, unlike other biological processes, was not felt to be a subject that is readily dissected into its components. Several laboratories have, however, spent the past decade chipping away at this notion. Although their collective efforts have not yet produced a completed work of art, they at least provide the first discernible outline of the molecular pathways that regulate lifespan.

Simple organisms

Yeast would seem an unlikely model system for learning about human ageing. A single-celled organism, essential perhaps for our daily bread and our favourite brew, it seems to lack the complexity that we associate with higher-order functions such as ageing. Nonetheless, yeast 'replicative lifespan' — a measure of the number of divisions a mother yeast cell can undergo — has become a useful surrogate for mammalian ageing. Early in life, a mother yeast cell can readily divide, asymmetrically, to produce a daughter cell. Later on, when the mother cell has divided many times, it begins to enlarge and its capacity to produce progeny diminishes. The fact that middle-aged yeast as well as middle-aged humans are generally slower, fatter and

less interested in reproduction provides our first (albeit broad) clue that certain biological markers of ageing — and so, perhaps, the underlying mechanisms — might have been conserved during evolution.

So what determines how many times a mother yeast cell can divide? Numerous environmental and genetic determinants have been identified. Among the environmental factors are several non-lethal stresses, for example reducing the glucose levels in the growth medium from 2% to 0.5%; such 'caloric restriction' can significantly increase the replicative lifespan of yeast. This concept of 'less is more' when it comes to calories and longevity is a theme that we see again and again, from yeast to mice.

In yeast, low-glucose conditions extend lifespan through the action of a gene termed *SIR2* (ref. 1). Strikingly, a similar lengthening of lifespan ensues when yeast are simply manipulated to produce too much of the protein product of this gene². This product, Sir2, was first identified as a protein that modifies the physical state of DNA, causing a phenomenon known as genetic silencing ('Sir' stands for 'silent information regulator'). Most evidence now supports the idea

Box 1 How long will you live?

The theoretical maximum lifespan in humans is still a subject of considerable debate. For most of us, however, our life expectancy is determined not by theoretical limits but by a confluence of factors we can't control (genetic) and factors we struggle to control (behavioural and environmental). For a hint on how long you might live — or at least how long an insurance company thinks you might live — simply fill out the questionnaire below. (From ref. 36.) Begin with the number 76 and add or subtract as follows:

If you are age 30–50 (+2), if 51–70 (+4) _____

Male (–3), Female (+4) _____

Live in an urban area of population more than 2 million (–2), or a town under 10,000 (+2) _____

One grandparent lived to age 85 (+2), or all grandparents lived to 80 (+6) _____

A parent died of cardiovascular disease or a stroke before age 50 (–4) _____

Brother, sister or parent under 50 has cancer or

heart disease, or diabetes since childhood (–3) _____

Earn the equivalent of over US\$50,000 (–2) _____

Finished college (+1), or have a graduate/professional degree (+2) _____

65 or over and still working (+3) _____

Live with a spouse or friend (+5) _____

Live alone (–3) and (–3) for every decade you lived alone since age 25 _____

Work behind a desk (–3), do work requiring strenuous physical labour (+3) _____

Exercise strenuously for 30 minutes 5 times a week (+4), 2–3 times a week (+2) _____

Sleep more than ten hours a day (–4) _____

Personality is relaxed (+3) or intense (–3), happy (+1) or unhappy (–2) _____

Speeding ticket in the previous year (–1) _____

Drink the equivalent of 1 oz of liquor per day (–1) _____



Smoke more than 2 packs per day (–8), 1–2 packs/day (–6), or 1/2–1 pack/day (–3) _____

Overweight by 50 lbs or more (–8), 30–49 lbs (–4), 10–29 lbs (–2) _____

If over age 40, do you have a check-up or (if female) see a gynaecologist annually? (+2) _____

Calculated life expectancy _____

M. S. YAMASHITA/CORBIS

that Sir2 extends lifespan in yeast by regulating gene expression or suppressing recombination (the exchange of chunks of DNA between chromosomes), although the relevant genetic targets of Sir2 are unknown.

But what is the link between caloric restriction and Sir2? The protein's effects on DNA are achieved through its histone deacetylase activity — its ability to remove specific acetyl groups from histones and other proteins that wrap up DNA. This activity of Sir2 in turn depends on the cellular levels of nicotinamide adenine dinucleotide (NAD)². NAD and its reduced form, NADH, represent a sort of basic energy currency in cells. Caloric restriction in yeast might increase Sir2 activity by altering either the NAD:NADH ratio or the levels of the NAD derivative nicotinamide. In other simple organisms, such as the fruitfly *Drosophila melanogaster*, caloric restriction might not only increase the activity of Sir2, but also directly regulate its levels³. Increased Sir2 levels and activity might then dampen gene expression and recombination, leading (somehow) to an extension in lifespan.

Too much Sir2 also extends lifespan in the nematode worm *Caenorhabditis elegans*². Like yeast, *C. elegans* has become a useful model in ageing research. Old and young worms look and act differently (Fig. 1), and the creature's lifespan is an experimentally manageable 20–25 days. Moreover, the worm genome has been completely sequenced, it is reasonably straightforward to mutate its genes and study the effects, and the expression of its genes can be easily manipulated using RNA interference (RNAi) methods—a technique for knocking out the middle step in the 'DNA to messenger RNA to protein' cascade that constitutes gene expression.

Over the past decade, studies of worms carrying mutations in one gene or another have revealed many genes besides *SIR2* that can, individually, regulate *C. elegans* lifespan (reviewed in ref. 4). We don't have space to describe them all here, but many seem to regulate metabolic parameters. Some of the longer-lived mutant worms, for instance, have difficulty eating and thus are caloric-restricted through genetic hardwiring. Another long-lived mutant, named *clk-1*, cannot synthesize coenzyme Q, a molecule that is involved in metabolism in mitochondria (the cell's major energy-generating bodies).

Finally, an interesting cluster of mutations affects the well-characterized insulin/insulin-like growth factor-1 (IGF-1) pathway. This signalling cascade is involved in nutritional sensing and metabolic regulation in a wide spectrum of organisms⁴. In long-lived animals with mutations in genes encoding components of this pathway, the activity of the DAF-16 protein (a member of the so-called Forkhead family, which regulates gene expression) is modestly increased (see Fig. 2 for why this happens). Two research groups^{5,6} have begun to track down the transcriptional



Figure 1 Life and death of a worm. **a**, Confocal images of the nematode worm *Caenorhabditis elegans*, showing the relative size and appearance of what might be viewed as, clockwise from top left, a baby, teenager and adult worm. Overlaying the images is the level of autofluorescence that the worm exhibits at these stages. This is believed to originate from the accumulation of oxidatively modified cellular components. **b**, Typical survival curves of worm populations. Maximum lifespan is denoted by green arrows; the mortality rate is indicated with blue lines and represents the slope of the curves. Genes that, when mutated or inactivated, extend lifespan shift the curve to the right (red arrow), potentially influencing the mortality rate as well.

targets of DAF-16. Although their different approaches uncovered non-overlapping sets of targets, both groups identified genes that apparently regulate metabolism and the organism's response to 'oxidative stress' — caused by the imbalance between intracellular levels of reactive oxygen species (also termed free radicals or oxidants), generated in part as a by-product of metabolism, and the cellular antioxidant system. The mammalian counterparts of DAF-16 activate the transcription of similar stress-related genes^{7,8}.

Studies of mutations in single genes are an effective but narrowly focused way of investigating the genetic basis of ageing. In the hope of defining the full spectrum of genes that regulate ageing, one group⁹ has turned to RNAi—a higher-throughput approach—to selectively inactivate more than 5,000 worm genes in turn. A familiar theme emerged: the most frequent category of gene product that appeared to extend lifespan when knocked out comprised regulators of mitochondrial function⁹. RNAi that was directed towards several different elements of mitochondrial electron transport, a crucial step in metabolism, also increased lifespan¹⁰ (although, curiously, only when mitochondrial function was inhibited in young worms).

From these studies, preliminary estimates suggest that, on average, inactivation of 1 out of every 50 random genes results in a reproducible lifespan extension⁹. In other words, these genes usually decrease lifespan. But why do worms — and presumably other species — hang on to so many life-reducing genes? One answer may be that, as our genes are (or so it is presumed) selected for only on the basis of their ability to promote reproduction and hence the continuance of the species, life-shortening genes must provide certain reproductive advantages to a young animal. In other words, as Charles Darwin proposed, and as those of us with teenage children can readily verify, post-reproductive adults are of little to no utility.

Single cells

Researchers have also been studying ageing by looking at mammalian cells in culture dishes. Like yeast, normal, non-reproductive mammalian cells generally lose their ability to proliferate after a fixed number of doublings—a phenomenon termed the Hayflick limit. With each ensuing division, a greater proportion of the cells enter a state called senescence, characterized by a permanent withdrawal from the cell-division cycle.

So what determines the Hayflick limit? One explanation for human cells is that every population doubling results in a shortening of the ends of each chromosome—protective caps called telomeres. Reduction of telomere length below a critical threshold can trigger senescence. Notably, normal non-reproductive cells can be 'immortalized' — meaning that they multiply indefinitely — when engineered to express telomerase, an enzyme that restores eroded telomeric DNA¹¹. The immortal nature of cancer cells, and of normal stem cells in embryos and adults, might relate to their continuous telomerase activity. Given these observations, some have argued that the forced expression of telomerase could potentially reverse or delay the ravages of ageing. Sceptics have warned, however, that this strategy might lead to unacceptable levels of malignancies — a case of the ends, telomeric or otherwise, not justifying the means.

Other manipulations of cells in culture influence lifespan, too. For instance, reduced levels of ambient oxygen delay senescence. The RAS gene, which regulates yeast replicative lifespan, can also contribute to mammalian cell senescence. Mutant Ras proteins are closely linked to the transformation of immortalized cells into tumour cells, but, paradoxically, the effect of Ras in a normal cell background is to block cell division irreversibly¹². This has strengthened the concept that senescence, like apoptosis (programmed cell death), is a protective mechanism. Interestingly, in both yeast and mammalian cells, RAS induces the production of a high level of mitochondrial oxidants, and in human cells these oxidants are required to induce senescence^{13,14}.

Mammals

Although the progress made in simple organisms and single cells has been remarkable, much less is known about mammalian ageing. But numerous parallels are emerging. For instance, several genetically different strains of long-lived mice have been described that are characterized by a deficiency in secreted hormones such as growth hormone and IGF-1 (ref. 15). These mice are at least vaguely reminiscent of those long-lived worms that show reduced IGF-1 signalling. Strikingly, mice in which the insulin receptor is inactivated only in adipose (fat) cells also show a 20% increase in their lifespan¹⁶, as do mice in which the IGF-1 receptor is only partially inactivated¹⁷. The tissue-specific inactivation of the insulin receptor is reminiscent of studies in worms in which increased DAF-16 activity in just a few tissues increased the organism's lifespan¹⁸.

Another regulator of mammalian lifespan is the p66^{shc} protein; deletion of this signal-transduction intermediate in mice results in a roughly 30% increase in lifespan¹⁹. Like many of the long-lived mutant worms, these mice can survive oxidative stress better than their normal litter-mates. Moreover, cells from these animals have lower levels of oxidants^{7,20}. Notably, p66^{shc} seems to regulate the activity of the mammalian Forkhead-family member that is a counterpart of *C. elegans* DAF-16 (ref. 7). Earlier this year, it was shown that mammalian Forkhead proteins also interact, physically and functionally, with NAD-dependent deacetylases from the Sir2 family^{21,22}. For all their evolutionary distance, mammals and nematodes seem to have kept the same bag of tricks.

Microarray analyses of the genes that are expressed in young and old tissues provide another means of trying to understand ageing in mammals. These analyses look at the messenger RNAs (mRNAs) that are expressed, and initial studies revealed that, in this respect, young and old tissues are remarkably similar^{23,24}. Indeed, comparative analysis of the brain and skeletal muscle

from old and young mice showed that only some 1–2% of all mRNAs differ by more than twofold in their expression levels. Those of us who wish to resist a mandatory retirement could argue that, at the cellular level, we are at roughly 98% of our youthful effectiveness. Nonetheless, although only a handful of genes may differ, those genes are clustered into functional classes that regulate the response to oxidative stress, inflammation and overall metabolism. Strikingly, these age-related changes in gene expression did not occur when the animals' calorie intake was restricted^{23,24}.

Whether or not human ageing and longevity could be influenced by caloric restriction is unknown, and understandably there are few people willing to participate in the lifelong experiment that would be necessary to find out. But further insight into human ageing has come from the study of

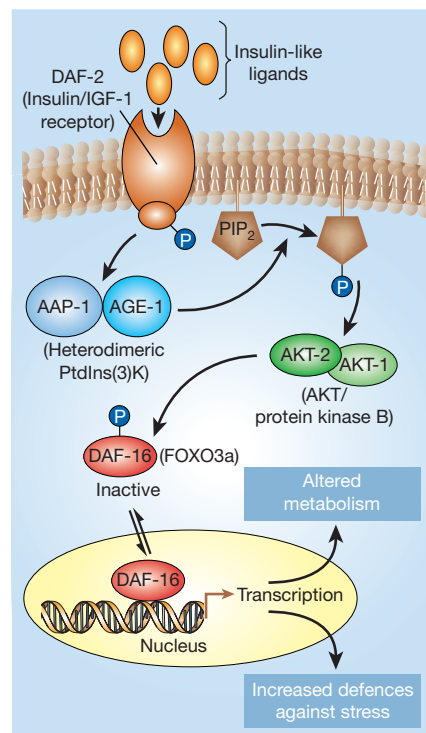


Figure 2 The DAF-16 signalling pathway and lifespan. The lifespan of the worm is regulated in part by a detailed biochemical pathway (simplified here) that influences the activity of the gene-transcription factor DAF-16. The binding of insulin-like molecules to their receptor (DAF-2) on the cell membrane activates a series of protein kinase enzymes: AAP-1–AGE-1 adds phosphate groups (P) to phosphatidylinositol-4,5-bis-phosphate (PIP₂), which activates AKT-1–AKT-2. This enzyme then phosphorylates DAF-16, keeping it in the cytoplasm. Some mutations that extend lifespan appear to modestly increase the nuclear localization and hence activity of DAF-16 in adult worms. The targets of DAF-16 include genes involved in overall metabolism or adaptation to stress. A related pathway exists in fruitflies. The closest mammalian counterparts of the key proteins are shown in parentheses.

the rare accelerated-ageing syndromes called progerias. The most widely known of these inherited disorders is Werner's syndrome, which is characterized by the early appearance of grey and thinning hair, osteoporosis, atherosclerosis and cancer²⁵. The gene that is altered in Werner's syndrome encodes a DNA helicase; other members of this enzyme family have been implicated in inherited predispositions to cancer. Such enzymes seem to be involved in normal DNA replication, repair and recombination.

These accelerated-ageing syndromes remind us of an important distinction between our understanding of ageing in people and that in lower organisms. For the most part, we die from specific diseases and events; the cause of death in lower organisms is less well understood. Can our understanding of the pathways that regulate ageing teach us anything about the mechanisms that underlie age-related diseases? If, for instance, we had a magic potion that slowed human ageing, would we also decrease the incidence of diseases such as atherosclerosis and cancer, which rises so dramatically as we get older? For the case of malignancies, some insight may come from a study in yeast showing that daughters derived from older mother yeast cells had a roughly 100-fold greater genetic instability than cells from younger mothers²⁶. If the same occurs in human cells, then cancer and ageing mechanisms may be inextricably intertwined.

In another example, circulating endothelial progenitor cells — which repair damaged blood vessels — decline as a function of human age²⁷. Similarly, in mice, the ability of resident skeletal-muscle progenitor cells to proliferate and hence repair damage appears to decline markedly as a function of age²⁸. These results hint that different tissues age at different rates, and that the exhaustion, depletion or senescence of progenitor cells (including stem cells) might contribute to overall ageing as well as age-related diseases. In this context, atherosclerosis, neurodegeneration and other specific disease states could be viewed as a form of tissue-specific progeria.

A unified theory?

During the 1992 US presidential campaign between Bill Clinton and George Bush Sr, out of a whole array of issues it was the economy that most exercised the electorate. In their campaign room the Clinton staff had a simple, and now infamous, sign to remind them to stay on message: "It's the economy, stupid!" In our view, the most likely sign to hang in any hypothetical ageing headquarters at the moment would read: "It's the free radicals, stupid!"

Indeed, the cellular effects of free radicals represent the most likely contender to explain the ageing process across a wide range of species. In a nutshell, the idea is that when mitochondrial metabolism — fuelled by food and oxygen — increases, more reactive

oxygen species are produced as a by-product, with greater adverse effects on cells.

Although such a theory was proposed close to 50 years ago²⁹, its current incarnation has nuances (Fig. 3). For instance, there is a growing appreciation that the production of reactive oxygen species in cells is intricately regulated, and that the effects of these species include both random damage to cellular components and the regulation of specific pathways³⁰. Similarly, although metabolic rate and the production of reactive oxygen species are generally correlated, this relationship can be complex. For instance, a group of mitochondrial proteins known as uncoupling proteins is activated by superoxide, one type of reactive oxygen species, and can induce a metabolic state in which oxygen consumption increases but overall levels of reactive oxygen species decline³¹. Conversely, expression of an activated RAS gene in yeast leads to decreased oxygen consumption but increased levels of reactive oxygen species¹³.

Nonetheless, many of the known regulators of longevity can be mapped to a scheme resembling Fig. 3. For instance, inhibiting mitochondrial function — as many of the genes uncovered in the *C. elegans* RNAi screens^{9,10} seem to do — might cause the organism to rely on alternative, non-mitochondrial pathways for energy production. These pathways are inherently less efficient, but produce fewer reactive oxygen species, perhaps leading to the lifespan extension observed. Mutations that activate DAF-16, meanwhile, would induce the expression of genes that encode antioxidants, and might also shift the organism to a different overall metabolic state. Caloric restriction could lead to many changes, including Sir2 activation and a consequent reduction in gene expression and recombination, as well as reduced activation of the insulin/IGF-1 pathways. The net effect might be to lower oxidant production and increase the removal of free radicals.

But although this theory is the best candidate to explain the wealth of experimental data, disquieting observations persist. Take two examples. First, if ageing represents the legacy of the combustible mixture of food and oxygen in our mitochondria, then oxidant-related damage should be, to some degree, cumulative. But when fruitflies were caloric-restricted for just two days, their mortality rate became equivalent to that of flies that spent their whole life hungry³².

Second, at least in yeast, glucose withdrawal extends life but results in a tripling in the rate of oxygen consumption³³. Oxygen consumption and the generation of reactive oxygen species are not necessarily equivalent. Still, one would expect such a large change in oxygen consumption to be more likely to increase than decrease the levels of such species. So the free-radical theory remains a useful, but still unproven, frame-

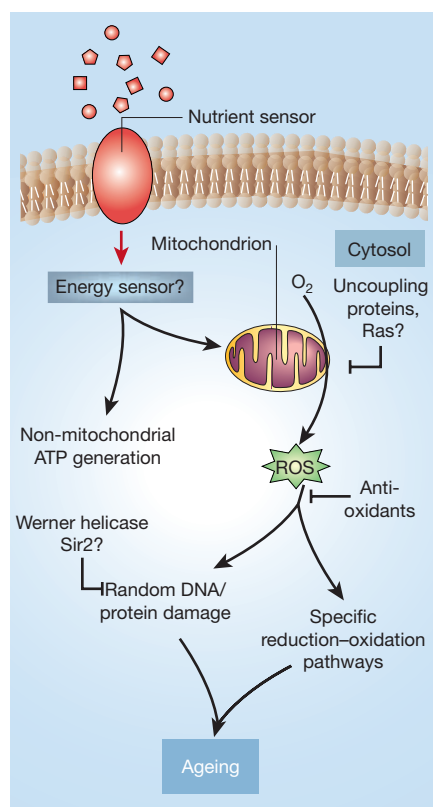


Figure 3 A current version of the free-radical theory of ageing. Nutrients (detected by sensors such as the insulin receptor at the cell membrane) and oxygen fuel metabolism in the mitochondria, generating reactive oxygen species (ROS). The levels of ROS might determine the rate of ageing by inducing random damage in proteins and DNA, and by activating specific pathways that depend on the reduction-oxidation status of particular proteins. DNA damage or recombination might be decreased by proteins that influence lifespan, including the Werner helicase or members of the Sir2 family. Although ROS levels and oxygen consumption are related, the relationship is complex; proteins such as uncoupling proteins and Ras may alter the levels of ROS produced or released per molecule of oxygen consumed. It also remains unclear how the cell partitions substrates to the inefficient but ROS-sparing non-mitochondrial pathways, or the highly efficient but ROS-generating mitochondrial pathways.

work within which to understand how we age, and questions remain as to whether levels of oxygen radicals cause, or are merely correlated with, ageing³⁴.

Conclusions

The discovery of genes that are associated with the regulation of lifespan in simple organisms, the apparent clustering of these genes into a handful of discrete categories and pathways, and the evolutionary conservation of such genes and pathways have led to the hope that treatments can be developed to slow ageing and increase lifespan — if we want to. The idea has obvious social

implications, with ageing populations already becoming a problem in many countries. Nonetheless, as caloric restriction seems to be such a robust way to extend life, there is considerable enthusiasm for the idea of molecules that mimic its effects. At least in yeast, such a strategy may hold water: last year, researchers identified numerous structurally related small molecules that increase Sir2 activity and so mimic the effects of caloric restriction on lifespan³⁵. For those of us who enjoy life's desserts, the fact that yeast exposed to these compounds lived longer despite being cultured in a food-rich environment provides some cause for optimism.

Other potentially promising compounds could include partial inhibitors of the insulin/IGF-1 pathways, small molecules that block p66^{shc}, or compounds that modestly activate the Forkhead-family proteins. Will such approaches be successful? Can we produce drugs that preserve the quality of life and yet extend its length? Some would argue that the proof that such strategies will work is already discernible from the menagerie of long-lived worms, fruitflies and mice in existence. Others would say that for evidence of how much remains a mystery, one need only visit a small graveyard in Arles and find a stone that bears the simple inscription: Jeanne Calment, born 1875, died 1997.

Shino Nemoto and Toren Finkel are in the Cardiovascular Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Building 10/6N-240, 10 Center Drive, Bethesda, Maryland 20892-1622, USA.

e-mail: finkelt@nhlbi.nih.gov

- Lin, S. J. *et al.* *Science* **289**, 2126–2128 (2000).
- Hekimi, S. & Guarente, L. *Science* **299**, 1351–1353 (2003).
- Rogina, B., Helfand, S. L. & Frankel, S. *Science* **298**, 1745 (2002).
- Murakami, S. *et al.* *Ann. NY Acad. Sci.* **908**, 40–49 (2000).
- Lee, S. S. *et al.* *Science* **300**, 644–647 (2003).
- Murphy, C. T. *et al.* *Nature* **424**, 277–283 (2003).
- Nemoto, S. & Finkel, T. *Science* **295**, 2450–2452 (2002).
- Kops, G. J. *et al.* *Nature* **419**, 316–321 (2002).
- Lee, S. S. *et al.* *Nature Genet.* **33**, 40–48 (2003).
- Dillin, A. *et al.* *Science* **298**, 2398–2401 (2002).
- Bodnar, A. G. *et al.* *Science* **279**, 349–352 (1998).
- Serrano, M. *et al.* *Cell* **88**, 593–602 (1997).
- Hlavata, L. *et al.* *EMBO J.* **22**, 3337–3345 (2003).
- Lee, A. C. *et al.* *J. Biol. Chem.* **274**, 7936–7940 (1999).
- Longo, V. D. & Finch, C. E. *Science* **299**, 1342–1346 (2003).
- Bluhm, M., Kahn, B. B. & Kahn, C. R. *Science* **299**, 572–574 (2003).
- Holzenberger, M. *et al.* *Nature* **421**, 182–187 (2003).
- Libina, N., Berman, J. R. & Kenyon, C. *Cell* **115**, 489–502 (2003).
- Migliaccio, E. *et al.* *Nature* **402**, 309–313 (1999).
- Trinei, M. *et al.* *Oncogene* **30**, 3872–3878 (2002).
- Motta, M. C. *et al.* *Cell* **116**, 551–563 (2004).
- Brunet, A. *et al.* *Science* **303**, 2011–2015 (2004).
- Lee, C. K. *et al.* *Science* **285**, 1390–1393 (1999).
- Lee, C. K. *et al.* *Nature Genet.* **25**, 294–297 (2000).
- Martin, G. M. & Oshima, J. *Nature* **408**, 263–266 (2000).
- McMurray, M. A. & Gottschling, D. E. *Science* **301**, 1908–1911 (2003).
- Hill, J. M. *et al.* *N. Engl. J. Med.* **348**, 593–600 (2003).
- Conboy, I. M. *et al.* *Science* **302**, 1575–1577 (2003).
- Harman, D. *J. Gerontol.* **2**, 298–300 (1957).
- Finkel, T. *Curr. Opin. Cell Biol.* **15**, 247–254 (2003).
- Echay, K. S. *et al.* *Nature* **415**, 96–99 (2002).
- Mair, W. *et al.* *Science* **301**, 1731–1733 (2003).
- Lin, S. J. *et al.* *Nature* **418**, 344–348 (2002).
- Sohal, R. S., Mockett, R. J. & Orr, W. C. *Free Rad. Biol. Med.* **33**, 575–586 (2002).
- Howitz, K. T. *et al.* *Nature* **425**, 191–196 (2003).
- www.gfcw.com/download.htm

Hyperactive antifreeze protein in a fish

This plasma protein offers the winter flounder extra protection against icy polar waters.

Fish that live in the polar oceans survive at low temperatures by virtue of 'antifreeze' plasma proteins¹ in the blood that bind to ice crystals and prevent these from growing. However, the antifreeze proteins isolated so far from the winter flounder (*Pleuronectes americanus*), a common fish in the Northern Hemisphere, are not sufficiently active to protect it from freezing in icy sea water. Here we describe a previously undiscovered antifreeze protein from this flounder that is extremely active (as effective as those found in insects) and which explains the resistance of this fish to freezing in polar and subpolar waters.

An antifreeze protein (AFP) known as type I was discovered in winter flounder 30 years ago² and has been extensively characterized^{3,4} as an alanine-rich, amphipathic molecule that is a single α -helix with a relative molecular mass (M_r) of 3,300 and binds to a pyramidal plane of ice⁵. The concentration of type I AFP found in the fish's blood plasma during winter is about 10–15 mg ml⁻¹ (ref. 6) and should result in a non-colligative freezing-point depression, or thermal hysteresis¹, of 0.7 °C (Fig. 1a). This thermal hysteresis, coupled with the freezing-point depression attributable to the colligative effect of blood solutes (0.8 °C)⁷, would protect the flounder down to a temperature of only -1.5 °C, which is critically short of what is required to survive at -1.9 °C, the freezing point of sea water. What mechanism could account for the additional 0.4 °C of protection against freezing?

To answer this question, we re-examined the thermal hysteresis activity of winter-flounder plasma and found it to be in excess of 2 °C. This is more than double the capability of the known type I AFP (Fig. 1a) and, when added to the colligative effect of plasma solutes, is more than enough to protect the fish below the freezing point of sea water. Lemon-shaped ice crystals that differ significantly from the hexagonal bipyramids obtained with pure type I AFP (Fig. 1b) indicate the presence of an unknown antifreeze protein.

The relative contribution of the two winter-flounder proteins to the antifreeze activity in the fish's plasma is illustrated by size-exclusion chromatography (Fig. 1b). Here the antifreeze activity in column-eluate fractions was measured as thermal hysteresis (total freezing-point depression less the colligative freezing-point depression of the buffer). The combined thermal hysteresis of fractions containing the larger protein (peak X in Fig. 1b) accounts for two-thirds of the total plasma thermal

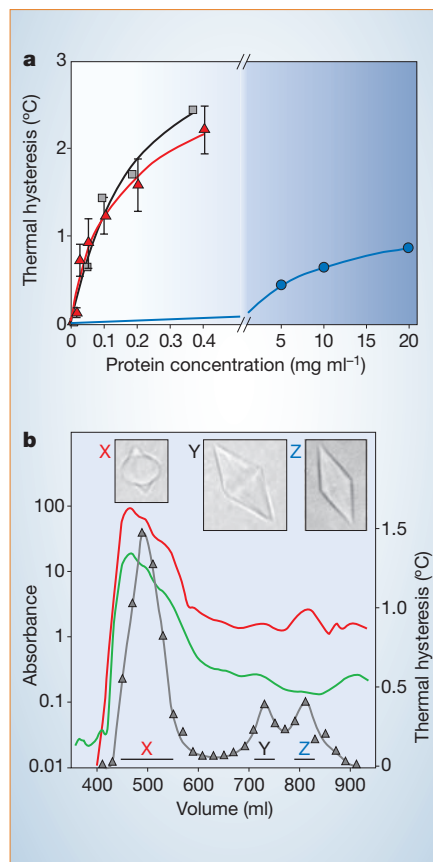


Figure 1 Fractionation and specific activity of winter-flounder antifreeze proteins (AFPs). **a**, Thermal hysteresis activity as a function of the concentration of AFP: type I (winter-flounder isoform HPLC-6 (ref. 10), blue circles); insect (yellow-mealworm isoform 4–9, black squares); 5a-like protein, with relative molecular mass 16,683, isolated from winter-flounder plasma (red triangles). **b**, Fractionation of winter-flounder plasma by gel-permeation chromatography on Sephadex G-75. The absorbance at 230 nm (red) and at 280 nm (green), and thermal hysteresis activity (black triangles) are shown for fractions across the chromatogram. Insets: morphology of the ice crystals produced from peak fractions X, Y and Z. MALDI-TOF mass spectrometry was used to identify the AFPs present in each peak: peak X contains the 5a-like AFP ($[M+H]^+$ average = 16,683.04); peak Y contains type-I isoform AFP-9 ($[M+H]^+$ monoisotopic = 4,282.24); peak Z contains type-I isoforms HPLC-6 and HPLC-8 ($[M+H]^+$ monoisotopic values of 3,239.98 and 3,226.01, respectively).

hysteresis, whereas the remainder (peaks Y and Z) is attributable to smaller proteins.

The size-separated antifreeze protein activities were associated with very different ice-crystal morphology (Fig. 1b, insets). The larger protein produced the lemon-shaped crystals seen in unfractionated plasma, whereas the smaller proteins produced hexagonal bipyramidal crystals typical of type I AFP. The identification of the type I AFP with M_r of 3,300, and of the less abundant one with M_r of 4,300, was confirmed by reversed-phase

high-performance liquid chromatography and mass spectrometry (Fig. 1b legend).

We further purified the larger AFP by anion-exchange chromatography and three rounds of ice-affinity purification⁸. The M_r of the resulting AFP was 16,683, which is comparable to, but distinct from, the M_r of 16,267 that is predicted for the exported product of a winter-flounder gene known as 5a (ref. 9). The two proteins differ slightly in their amino-terminal sequence and amino-acid composition. At the time of its discovery, the 5a gene was dismissed as an antifreeze-protein pseudogene, largely because the protein it encodes would have been grossly different from type I AFP and had never been detected in the flounder.

The newly discovered flounder protein that so closely resembles the 5a product is extraordinarily active in comparison with other fish antifreeze proteins (Fig. 1a). At a concentration of 0.1 mg ml⁻¹ it provides 1.1 °C of thermal hysteresis, equivalent to that of an insect antifreeze protein, whereas type I AFP and other fish antifreeze proteins at this concentration depress the freezing point by less than 0.1 °C.

This protein has remained undetected for 30 years because it is extremely labile and present in only moderate quantities. It irreversibly loses all activity at room temperature and at low pH — conditions previously used in purifying antifreeze proteins². The circulating concentration of this highly active protein is about 0.2 mg ml⁻¹, which is some 50-fold less than that of type I AFP.

The evolutionary relationship between our 5a-like antifreeze protein and type I AFP, which also contains short tracts of alanine, remains to be solved.

**Christopher B. Marshall*,
Garth L. Fletcher†, Peter L. Davies***

*Department of Biochemistry, Queen's University,
Kingston, Ontario K7L 3N6, Canada
e-mail: daviessp@post.queensu.ca

†Ocean Sciences Centre, Memorial University of
Newfoundland, St John's, Newfoundland A1C 5S7,
Canada

1. Fletcher, G. L., Hew, C. L. & Davies, P. L. *Annu. Rev. Physiol.* **63**, 359–390 (2001).
2. Duman, J. G. & DeVries, A. L. *Nature* **274**, 237–238 (1974).
3. Yang, D. S., Sax, M., Chakrabarty, A. & Hew, C. L. *Nature* **333**, 232–237 (1988).
4. Sicheri, F. & Yang, D. S. *Nature* **375**, 427–431 (1995).
5. Knight, C. A., Cheng, C. C. & DeVries, A. L. *Biophys. J.* **59**, 409–418 (1991).
6. Slaughter, D. & Hew, C. L. *Can. J. Biochem.* **60**, 824–829 (1982).
7. Scott, G. K., Davies, P. L., Kao, M. H. & Fletcher, G. L. *J. Mol. Evol.* **27**, 29–35 (1988).
8. Kuiper, M. J., Lankin, C., Gauthier, S. Y., Walker, V. K. & Davies, P. L. *Biochem. Biophys. Res. Commun.* **300**, 645–648 (2003).
9. Davies, P. L. & Gauthier, S. Y. *Gene* **112**, 171–178 (1992).
10. Hew, C. L. *et al. Eur. J. Biochem.* **160**, 267–272 (1986).

Competing financial interests: declared none.

Reproductive biology

Delivering spermatozoan RNA to the oocyte

Even though the genetic fingerprint of human sperm has been defined, its role in orchestrating fertilization and the development of the early embryo remains vague. Here we show that human male gametes pass over more to the oocyte than just the haploid male genome — paternal messenger RNAs are also delivered to the egg at fertilization. If these transcripts, previously thought to be left-overs from spermatogenesis, are important in early development, our findings may have implications for the success of somatic-cell nuclear transfer in cloning technology and the identification of components leading to unexplained male-factor infertility.

Microarray analysis has identified a set of transcripts that are common to both human testes and spermatozoa and are implicated in fertilization as well as in zygotic and embryonic development¹. We investigated this possibility using the polymerase chain reaction (PCR) with reverse transcription to determine which of these transcripts is present in human spermatozoa but not in unfertilized human oocytes, and identified six candidates (Table 1). The implication is that spermatozoa deliver these transcripts to the ooplasm at fertilization.

We therefore used the zona-free hamster egg/human sperm penetration assay (a standard clinical assay to test the virility of human sperm) to investigate this possibility (for methods, see supplementary information). Complementary DNAs from human spermatozoa, hamster oocytes and hybrid human–hamster zygotes were interrogated by real-time PCR; these RNA preparations were devoid of DNA contamination (the 310-base-pair protamine-2 gene was not detected). Only the protamine-2 and clusterin transcripts were consistently detected in spermatozoa, zygotes and in the positive control, and not detected in hamster oocytes or in the negative control (Fig. 1). These results show that spermatozoa deliver

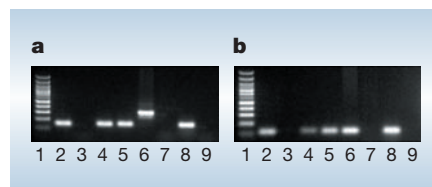


Figure 1 Early fate of two human spermatozoa-gene transcripts delivered at fertilization. **a**, Protamine-2; **b**, clusterin. All RNAs, including messenger RNAs, were isolated from human spermatozoa, hamster oocytes and zygotes of hamster oocytes fertilized by human spermatozoa at 30 min and 3 h post-fertilization. The transfer of spermatozoa-specific RNA was assayed by real-time polymerase chain reaction (PCR): half of the reaction mixture was used to confirm product size on 2% agarose gels stained with ethidium bromide. PCR fidelity was verified with human genomic DNA and testis complementary DNA (cDNA) as positive controls and hamster genomic DNA and water as negative controls. Lanes in both gels: (1) size-calibration 'ladder' (100–1,000 base pairs (bp) at 100-bp increments, plus 1,200- and 1,500-bp standards); (2) human spermatozoa cDNA; (3) hamster oocyte cDNA; (4) zygotic cDNA at 30 min post-fertilization; (5) zygotic cDNA at 3 h post-fertilization; (6) human genomic DNA; (7) hamster genomic DNA; (8) human testis cDNA; (9) water. The protamine-2 sequences (148 bp for cDNA; 310 bp for genomic DNA) or clusterin sequences (110 bp) are evident in spermatozoa, zygotes and the positive control, but not in hamster oocytes or in the negative control.

RNAs to the oocyte at fertilization.

Why should spermatozoa messenger RNAs be transferred to the oocyte? Messenger RNAs encoding proteins that bind nucleic acids, such as protamine-2, are likely to be deleterious and are probably degraded following entry², and a similar fate may await other RNAs that gain access³. But some may have a role in the developing zygote: for example, clusterin (also known as sulphated glycoprotein-2, or SGP-2) is delivered to the oocyte and has been implicated in cell–cell and cell–substratum interactions, enhancement of fertility rate, lipid transportation, membrane recycling, stabilization of stress proteins, and promotion or inhibition of apoptosis^{4–7}. These may therefore be required in the early zygote but unnecessary in the oocyte. Alternatively (or in addition), these and other unidentified molecules, such as small interfering RNAs (siRNAs), may participate in processes such as pro-

development up to the limb-bud stage (corresponding to about 30 somites)^{8,9}, and only recently a 'parthenogenetic' (derived from two different egg cells) mammal was born after genome modification¹⁰. However, the success of such experiments and of somatic-cell nuclear transfer is limited¹¹, as is the production of human embryonic stem cells after somatic-cell nuclear transfer¹². This may be because sperm RNAs contribute to early development. Transcripts that are specific to male germ cells play a role in the differentiation of embryonic stem cells¹³ and their function may not be easily replaced.

Our results indicate that sperm RNAs could be important in early zygotic and embryonic development and that they may hold the key to more successful somatic-cell nuclear transfer or to the identification of male-derived factors that underlie idiopathic infertility.

G. Charles Ostermeier*†, David Miller‡, John D. Huntriss‡, Michael P. Diamond*, Stephen A. Krawetz*†§

*Department of Obstetrics and Gynecology, †Center for Molecular Medicine and Genetics, §Center for Scientific Computing, Wayne State University, Detroit, Michigan 48201, USA

e-mail: steve@compbio.med.wayne.edu

‡Reproductive and Early Development Group, University of Leeds, Leeds LS2 9JT, UK

- Ostermeier, G. C., Dix, D. J., Miller, D., Khatri, P. & Krawetz, S. A. *Lancet* **360**, 772–777 (2002).
- Ziyyat, A. & Lefevre, A. *Hum. Reprod.* **16**, 1449–1456 (2001).
- Hayashi, S., Yang, J., Christenson, L., Yanagimachi, R. & Hecht, N. B. *Biol. Reprod.* **69**, 1170–1176 (2003).
- O'Bryan, M. K. et al. *J. Clin. Invest.* **85**, 1477–1486 (1990).
- Poon, S. et al. *J. Biol. Chem.* **277**, 39532–39540 (2002).
- Trougakos, I. P. & Gonos, E. S. *Int. J. Biochem. Cell Biol.* **34**, 1430–1448 (2002).
- Wong, P. et al. *Eur. J. Biochem.* **221**, 917–925 (1994).
- Kono, T., Sotomaru, Y., Katsuzawa, Y. & Dandolo, L. *Dev. Biol.* **243**, 294–300 (2002).
- Kure-Bayashi, S., Miyake, M., Okada, K. & Kato, S. *Theriogenology* **53**, 1105–1119 (2000).
- Kono, T. et al. *Nature* **428**, 860–864 (2004).
- Wilmot, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. *Nature* **385**, 810–813 (1997).
- Hwang, W. S. et al. *Science* **303**, 1669–1674 (2004).
- Geijsen, N. et al. *Nature* **427**, 148–154 (2004).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Corrigendum

Sex differences in learning in chimpanzees

Elizabeth V. Lonsdorf, Lynn E. Eberly, Anne E. Pusey
Nature **428**, 715–716 (2004)

The x-axis of the lower panel of Fig. 1b should have been scaled as 0–100%, not 0–1% as published.

brief communications arising online

♦ www.nature.com/bca

Photonics: Tuning holes in photonic-crystal nanocavities

C. Sauvan, P. Lalanne & J.-P. Hugonin
(doi:10.1038/nature02602)

Reply: T. Asano & S. Noda (doi:10.1038/nature02603)

Table 1 Spermatozoa transcripts implicated in early development

Fertilization	Stress response	Embryogenesis and morphogenesis	Implantation
+ Clusterin	HSF2	MID1	–RPL2
Calnexin	HSPA1L	–NLVCF	
+ AKAP4	DNAJB1	CYR61	
– Oscillin	+ HSBP1(CDH13)	EYA3	
+ Protamine-2	DUSP5	+ FOXG1B	
		+ WNT5A	
		– WHSC1	
		– SOX13	

Transcripts detected in spermatozoa for biological processes important to fertilization and its associated stress response, which is critical for sustaining zygotic development and early embryonic development. Transcripts that are detected in spermatozoa and not in unfertilized human oocytes are indicated by plus and minus signs, respectively. We can roughly estimate the total number of RNA molecules that would be delivered to the oocyte upon fertilization as some 18,000 transcripts, assuming an average RNA size of 1,000 bases and a yield of RNA of at least 10 ng per million spermatozoa.

Photonics

Tuning holes in photonic-crystal nanocavities

Arising from: Y. Akahane, T. Asano, B.-S. Song & S. Noda *Nature* **425**, 944–947 (2003)

One challenge in photonics is strongly to confine light in small volumes in order to increase light–matter interaction. Akahane *et al.*¹ propose a new concept for increasing the lifetime of this interaction, based on tailoring of the Fourier spectrum of cavity modes, which they believe is demonstrated by the surprising enhancement (roughly tenfold) of the quality factor Q of the cavity as a result of fine-tuning the mirror-hole geometry in a photonic-crystal nanocavity. Here we question the validity of their concept and argue that the improvement in Q is due to an increase in the impedance wave matching at the cavity edges and to a slow-wave effect. This alternative interpretation opens the way to new cavity designs.

Like Akahane *et al.*¹, we first consider two cavities formed in a silicon slab with air claddings (Fig. 1a). The light is confined in the x -direction by perfect-metal mirrors (left cavity) or by Bragg mirrors (right cavity). For the cavity parameters given in the figure legend and for y -polarization, the resonant wavelength is $\lambda = 1.5 \mu\text{m}$. Using the Fourier modal method², we calculate the cavity mode profiles at the centre plane of the slab (Fig. 1b) and their Fourier spectra (Fig. 1c).

These curves closely resemble those in Fig. 2 of Akahane *et al.*, which would suggest¹ that the metallic cavity, with large Fourier components in the leaky region, should show a smaller Q factor than the dielectric one.

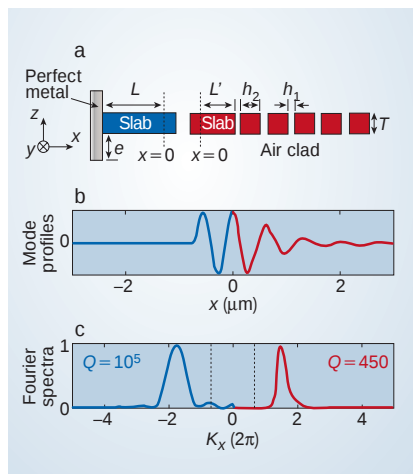


Figure 1 Counterexample for the Fourier space method. **a**, Cavity geometries. Parameters are: $T = 0.2 \mu\text{m}$, $e = 0.5 \mu\text{m}$, $2L' = 0.37 \mu\text{m}$, $2L = 1.38 \mu\text{m}$, $h_2 = 0.25 \mu\text{m}$, $h_1 = 0.1 \mu\text{m}$; the silicon refractive index is 3.4. As both cavities are symmetric with respect to the $x=0$ plane, only half of each is represented. **b**, Real part of the electric field profiles of the cavity mode at the centre plane of the slab. **c**, Modulus squared of the spatial Fourier spectra of the cavity modes. Vertical dashed lines delimitate the domain of leakage into the air clads, $|k_x| < 2\pi/\lambda$.

However, the metallic-cavity Q is 10^5 , a value that is 200 times larger than that of the dielectric cavity. In addition, for a null metal height of e , the mode profile of the metallic cavity remains completely unchanged but its Q is only 350.

We infer that the mode profile inside the cavity conveys nothing about the mode lifetime. The Fourier spectrum method is valid for calculating mode profiles across a surface positioned above the cavity in the air clad³. For such x -invariant surfaces, the cavity-mode profile and its Fourier spectrum contain all the information on the far-field radiation losses.

But this is not true for a surface inside the cavity. The reason is that the parallel k_x momentum of the cavity mode is not strictly matched at the slab–cladding interface because neither Snell’s law nor Fresnel’s law apply at a corrugated interface. Design rules that overlook this parallel-momentum mismatch ignore the essence of the longitudinal confinement and, in turn, the associated impedance-mismatch problem⁴ that determines radiation losses at the cavity edges.

To understand the physical reasons for the roughly tenfold enhancement in Q , we consider a Fabry–Perot model. We assume that the nanocavity mode is formed by the recirculation of the fundamental Bloch wave of the line-defect waveguide between the mirrors. The model, which relies on a three-dimensional computation² for the Bloch-wave modal reflectivity $|r_m|\exp(i\phi)$, reproduces the experimental trends well: a red-shift of λ as the hole shift d increases, and a peak value of Q for $d \approx 0.17a$, where a is the lattice constant of the photonic crystal. For Fabry–Perot cavities⁵,

$$Q = \frac{\pi}{1 - |r_m|^2} \left(\frac{2L}{\lambda} n_g - \frac{\lambda}{\pi} \frac{\delta\phi}{\delta\lambda} \right) \quad (1)$$

where $L = 3a$ is the defect length. From equation (1), three effects explain the Q improvement between $d = 0$ and $0.17a$: an increase of $|r_m|^2$ from 98.3% to 99.7%, a doubling of the group-index n_g of the Bloch wave, and a doubling of the penetration depth $\delta\phi/\delta\lambda$ into the mirrors.

The first effect has been analysed previously for a one-dimensional slab⁶ and two-dimensional air-bridge cavities⁷, and the second and third effects result from the highly dispersive nature^{8,9} of the Bloch wave in the spectral window. Our analysis highlights the importance of wave impedance matching in nanocavities and the key role of slow waves in further improvements.

Christophe Sauvan, Philippe Lalanne, Jean-Paul Hugonin

Laboratoire Charles Fabry de l’Institut Optique, Centre National de la Recherche Scientifique et Université Paris Sud, BP 147, 91 403 Orsay Cedex, France
e-mail: philippe.lalanne@iota.u-psud.fr
doi:10.1038/nature02602

1. Akahane, Y., Asano, T., Song, B.-S. & Noda, S. *Nature* **425**, 944–947 (2003).
2. Silberstein, E., Lalanne, P., Hugonin, J. P. & Cao, Q. *J. Opt. Soc. Am. A* **18**, 2865–2875 (2001).
3. Vuckovic, J., Loncar, M., Mabuchi, H. & Scherer, A. *IEEE J. Quant. Electr.* **38**, 850–856 (2002).
4. Palamaru, M. & Lalanne, P. *Appl. Phys. Lett.* **78**, 1466–1468 (2001).
5. Coldren, L. A. & Corzine, S. W. (eds) *Diode Lasers and Photonic Integrated Circuits* (Wiley, New York, 1995).
6. Lalanne, P. & Hugonin, J.-P. *IEEE J. Quant. Electr.* **39**, 1430–1438 (2003).
7. Lalanne, P., Mias, S. & Hugonin, J. P. *Opt. Exp.* **12**, 458–467 (2004).
8. Notomi, M. *et al.* *Phys. Rev. Lett.* **87**, 253902 (2001).
9. Chutinan, A. & Noda, S. *Phys. Rev. B* **62**, 4488–4492 (2000).

Akahane et al. reply — In response to Sauvan *et al.*¹, we perceive a flaw in their use of the Fourier analysis method and confirm that the increased Q factor in a nanocavity with tuned air holes is based on our design concept².

Referring to the cavity shown with a metal mirror in Fig. 1 of Sauvan *et al.*, the slab cavity is surrounded by free space (or air clad) in the vertical direction when $e = 0$, which corresponds to the case² in our Fig. 2b, c. The lifetime of the cavity mode is determined by the amount of its Fourier components inside the ‘leaky region’ determined by air clad. When e becomes non-zero, however, the situation changes drastically, because the outer space of the slab cavity changes from the air clad to a hollow cavity (or waveguide) that is surrounded by metals and the slab surface³. It means that the leaky region is no longer determined by the air clad but by the hollow cavity. Because the number of modes in the hollow cavity³ is much less than that of those in air clad, the leaky region is very much reduced. Thus, the leakage of light from the slab cavity to the outer space is decreased, and the Q factor increases.

Sauvan *et al.* overlook this change in the leaky region that is associated with the change in the outer space, which leads to a serious misunderstanding. The method itself (to increase the Q factor) is interesting, but in reality the metal mirror has considerable absorption and is hard to use for an optical regime. Thus, we concentrated on the Bragg reflector cavity type².

When we investigate the mode lifetime of the Bragg reflector cavity surrounded by air clad, we can apply a normal Fourier analysis method by using the leaky region determined by the air clad. We described an important design concept² in which the envelope of the in-plane mode profile in the cavity should be varied gently but localized spatially (gaussian-like, lorentzian-like or like the related function) to increase Q while keeping a small

modal volume, V , where the components inside the leaky region in the Fourier space are much reduced (Fig. 2d, e of ref. 2).

In the case of our two-dimensional photonic-crystal slab cavity, we showed that the in-plane mode profile approaches gaussian-like when the air holes are shifted appropriately. On the other hand, in the case of the Bragg cavity of Sauvan *et al.* (their Fig. 1), we have confirmed that the in-plane mode profile approaches lorentzian-like when the air holes are shifted appropriately. We have cal-

culated the in-plane mode profiles, both at the centre plane of the cavity and at the surface just above the cavity, and confirmed that both profiles are very similar and that the Fourier analyses on each give us enough hints on the mode lifetime. Our design concept therefore works well for both our cavities and theirs.

We question the discussion by Sauvan *et al.* that is based on their equation (1) in view of the fact that they do not provide any evidence that the change in the calculated

reflectivity r_m for Bragg reflectors, with and without shift of air holes, is due to an increase in impedance wave matching.

Takashi Asano, Susumu Noda

*Department of Electronic Science and Engineering,
Kyoto University, Kyoto 615-8510, Japan*

e-mail: snoda@kuee.kyoto-u.ac.jp

doi:10.1038/nature02603

1. Sauvan, C., Lalanne, P. & Hugonin, J.-P. *Nature* doi:10.1038/nature02602 (2004).
2. Akahane, Y., Asano, T., Song, B.-S. & Noda, S. *Nature* **425**, 944–947 (2003).
3. Waymouth, J. F. *J. Light Vis. Env.* **13**, 51–68 (1989).

The formation of a massive protostar through the disk accretion of gas

Rolf Chini¹, Vera Hoffmeister¹, Stefan Kimeswenger², Markus Nielbock¹, Dieter Nürnberg³, Linda Schmittobreck³ & Michael Sterzik²

¹Astronomisches Institut, Ruhr-Universität Bochum, 44780 Bochum, Germany

²Institut für Astrophysik, Universität Innsbruck, 6020 Innsbruck, Austria

³European Southern Observatory, Santiago 19, Chile

The formation of low-mass stars like our Sun can be explained by the gravitational collapse of a molecular cloud fragment into a protostellar core and the subsequent accretion of gas and dust from the surrounding interstellar medium^{1–3}. Theoretical considerations suggest that the radiation pressure from the protostar on the in-falling material may prevent the formation of stars above ten solar masses through this mechanism⁴, although some calculations have claimed that stars up to 40 solar masses can in principle be formed via accretion through a disk^{5–7}. Given this uncertainty and the fact that most massive stars are born in dense clusters, it was suggested that high-mass stars are the result of the runaway merging of intermediate-mass stars⁸. Here we report observations that clearly show a massive star being born from a large rotating accretion disk. The protostar has already assembled about 20 solar masses, and the accretion process is still going on. The gas reservoir of the circumstellar disk contains at least 100 solar masses of additional gas, providing sufficient fuel for substantial further growth of the forming star.

The Omega nebula (M 17) is one of the most prominent star-forming regions in our Galaxy⁹ at a distance of 2,200 parsecs (1 pc = 3.086 × 10¹⁸ cm). M 17 is extremely young, as shown by the presence of several high-mass stars which ionize the surrounding hydrogen gas¹⁰; the total luminosity of these stars exceeds that of our Sun by a factor of almost 10⁷. Adjacent to the southwestern edge of the H II region, there is a huge cloud of molecular gas, which is believed to be a site of ongoing star formation.

We have recently investigated this interface between the H II region and the molecular cloud by means of very deep infrared imaging (between 1.2 and 2.2 μm wavelength) in order to search for newly forming high-mass stars. Our sensitive measurements penetrated the southwestern molecular cloud so that the faint nebular emission of the H II region, which is partly located behind the molecular cloud, could shine through the dust. At right ascension (RA) 18 h 20 min 26.3 s, declination (dec.) –16° 12' 10" (J2000), we discovered a large opaque silhouette against the nebular background associated with an hourglass-shaped reflection nebula. As we will show below, this system complies perfectly with theoretical predictions for a newly forming high-mass star surrounded by a huge accretion disk and accompanied by an energetic bipolar mass outflow.

The most obvious morphological components of the system are two triangular-shaped dust lanes that are apparent at infrared wavelengths. Figure 1 shows a high-resolution 2.2 μm image, which suggests that the morphology of the silhouette resembles a flared disk, seen nearly edge-on. The disk has a diameter of about 20,000 astronomical units (1 AU = 1.496 × 10¹³ cm), which is 250 times the size of our Solar System and by far the largest circumstellar disk detected to date. Although there are a small number of candidates for massive young stellar objects (YSOs), some of which are associated with outflows^{11–14}, the largest circumstellar disk hitherto detected around these objects has a diameter of only 130 AU (ref. 15). NGC 7538S, another potential massive YSO and a site of OH and H₂O masers, is embedded in an elongated cloud core¹⁶. Molecular line studies indicate both rotation and outflow, but the velocity structure of the entire system is extremely confused

and there is no direct evidence for a disk. The largest silhouette disk so far is known as 114-426 in Orion and has a diameter of about 1,000 AU; however, its central star is probably a low-mass object rather than a massive protostar.

The ¹³CO data indicate that the disk/envelope system slowly rotates, with its northwestern part approaching the observer (Fig. 2). Over an extent of 30,800 AU we measure a velocity shift of 1.7 km s^{–1}, that is, a velocity gradient of about 11 km s^{–1} pc^{–1} (Fig. 3). Adopting an abundance ratio of 43 between ¹²CO and ¹³CO and a conversion factor of $N(\text{H}_2) = 2.3 \times 10^{20} \text{ cm}^{-2}$ per K km s^{–1} to derive H₂ column densities from the velocity integrated ¹²CO intensities, we obtain a conservative lower limit for the disk mass of 110 solar masses. Several effects will increase the true disk mass: (1) all observed lines are more (¹²CO) or less (¹³CO, C¹⁸O) optically thick. (2) The CO lines may be partly photodissociated. (3) The above conversion factor is more compatible with regions of low densities ($n_{\text{H}} \approx 200 \text{ cm}^{-3}$) and low temperatures ($T \approx 10 \text{ K}$) and thus may not be applicable in a hot dense core environment of a massive star, or in an externally heated and compressed cloud edge such as the M 17 ionization front. Given that the conversion factor varies with $n^{0.5}/T$, the disk mass would, for example, increase by a factor of three when using $n_{\text{H}} \approx 10^5 \text{ cm}^{-3}$, as suggested by the column densities from the optical depth at 2.2 μm and assuming a temperature of $T \approx 100 \text{ K}$.

The second morphological structure that is visible throughout the entire spectral range from 0.4 to 2.2 μm is an hourglass-shaped nebula perpendicular to the plane of the disk; Fig. 4 shows a sequence of images. Its spectrum is dominated by the emission lines of Hα, the Ca II triplet 8,498, 8,542 and 8,662 Å, and He I 6,678 Å (Fig. 5). In the case of low-mass stars, these lines provide indirect evidence for ongoing accretion from the inner disk onto the star^{17,18}. The Ca II triplet was also shown to be a product of disk accretion for both a large sample of T Tauri and Herbig Ae/Be stars¹⁹. The Hα line is extremely broad and shows a deep blue-shifted absorption as well as an inverse P Cygni profile. Blue-shifted

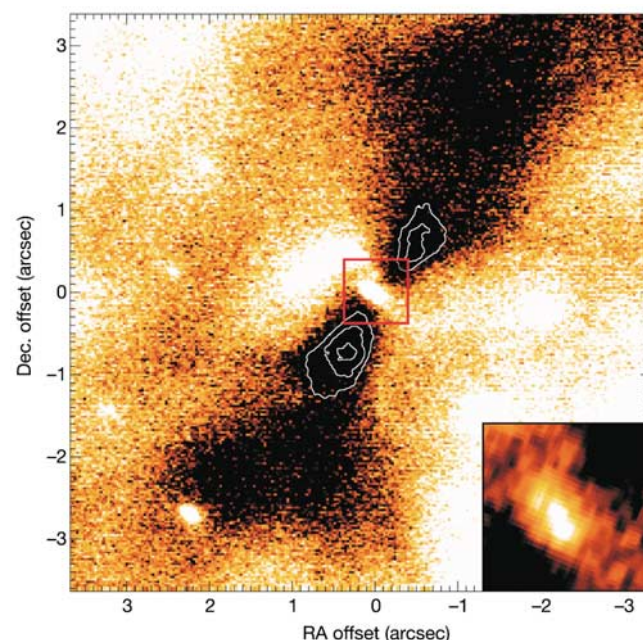


Figure 1 Silhouette of the accretion disk seen against the light from the H II region. The 2.2 μm image is centred at RA 18 h 20 min 26.3 s, dec. –16° 12' 10" (J2000), tick marks denote 1" and the effective spatial resolution is 0.25". White contours delineate the densest part of the disk (inner torus). The visible stars (slightly elongated owing to the adaptive optics technique) are embedded within the molecular cloud, but are probably unrelated to the disk. Inset, a deconvolved zoomed version of the central object of about 450 × 240 AU; its major axis is tilted by ~15° against the direction of the disk axis.

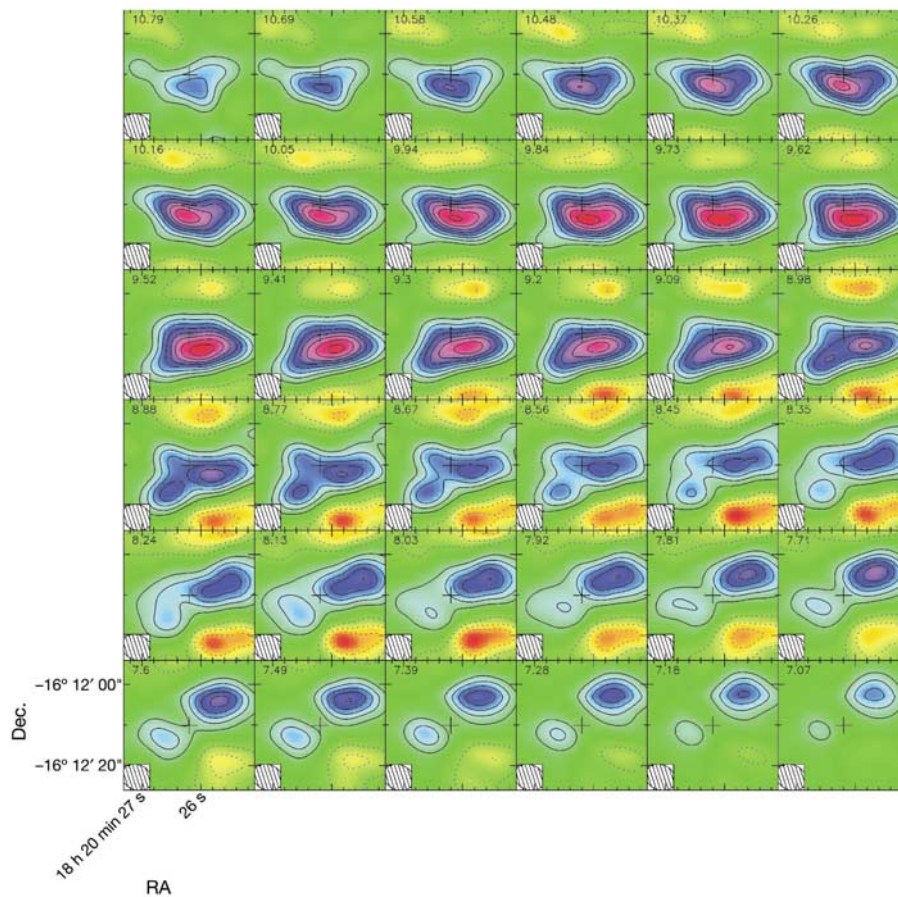


Figure 2 Kinematics of the molecular gas in the disk region. The ^{13}CO (1–0) channel maps of the accretion disk as observed with the Plateau de Bure interferometer show that its orientation is almost edge-on, with the major axis running from southeast to northwest. However, owing to the limited size of the synthesized beam ($7.8'' \times 6.2''$, $\text{PA} \approx 63^\circ$) and

the confusion by emission from outflowing and infalling material, the kinematics of the inner part of the disk is quite complicated and not traced properly by ^{13}CO . The numbers in each small panel denote the velocity v_{LSR} of the corresponding emission with respect to the local standard of rest.

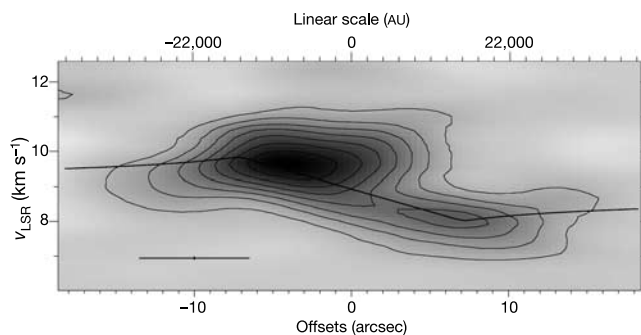


Figure 3 Position–velocity diagram showing the rotation of the disk. It is derived from a cut along the major axis of the disk, running at an angle of 45° across the ^{13}CO data cube displayed in Fig. 2. Greyscale (from -0.3 to 2.4 mJy per beam) and contours (at levels from 30% to 90% of the peak flux density of 2.4 mJy per beam) denote the intensity of the ^{13}CO emission. For comparison, we show the theoretically expected position–velocity curve for an edge-on disk around a 15-solar-mass star; the outer part (radii above $15,400$ AU) is in keplerian rotation, while its inner part is modelled as a rigid rotator. Error bars for both the angular ($\sim 7''$) and spectral (0.1 km s^{-1}) resolution of the ^{13}CO data are given in the lower left corner.

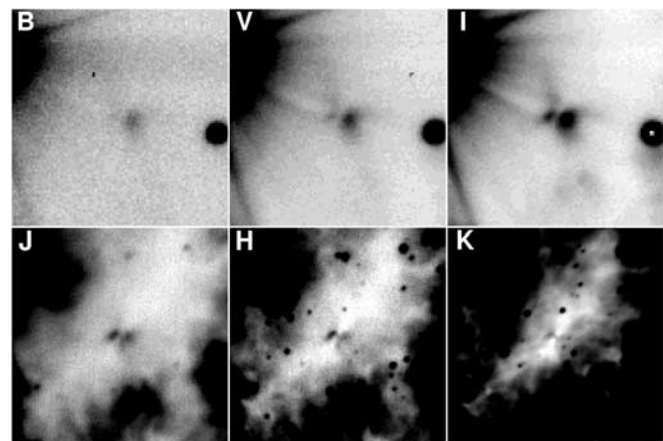


Figure 4 Multi-wavelength sequence of the hourglass-shaped reflection nebula. The central position of each frame is at RA $18 \text{ h } 20 \text{ min } 26.3 \text{ s}$, dec. $-16^\circ 12' 10''$ (J2000); the size is always $30'' \times 30''$. The change in morphology and intensity is delineated from the optical to the near infrared; the corresponding wavelengths are denoted at the top left of each panel (B, $0.45 \mu\text{m}$; V, $0.55 \mu\text{m}$; I, $0.8 \mu\text{m}$; J, $1.25 \mu\text{m}$; H, $1.6 \mu\text{m}$; and K, $2.2 \mu\text{m}$). In the upper row, the nebula is dominated by reflected light from a bipolar cavity; the southwestern (right) lobe is much stronger while the northeastern (left) lobe is obscured by the disk. The lower row is dominated by more compact emission (see also Fig. 1).

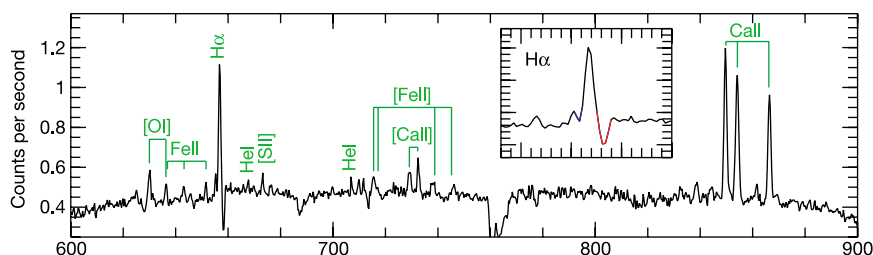


Figure 5 Optical spectrum of the bipolar nebula. The slits—oriented perpendicular to the disk plane—had lengths of 5' and 8' and a width of 1". The contamination from the H II region has been subtracted by averaging the emission from several offset positions. The x

axis gives the wavelength in nanometres. The various emission lines in the reflected light indicate evidence for ongoing accretion.

absorption components in permitted lines are typically associated with accretion disk-driven outflows^{20,21}; the same is true for forbidden emission lines such as [O I] 6,300 Å and [S II] 6,731 Å, which are also present in our spectrum. Numerous permitted and forbidden Fe II lines that are velocity-shifted by $\pm 120 \text{ km s}^{-1}$ are clear signs of high-velocity dissociative shocks with velocities of more than 50 km s^{-1} (ref. 22).

Owing to heavy extinction, the nature of an accreting protostellar object is difficult to infer. For all suggested massive YSOs (see examples above), there is no direct evidence for a (proto)stellar central object; likewise, the origin of the luminosity—typically about 10^4 solar luminosities—is unclear, and may be due to multiple objects or even embedded clusters. The new disk in M 17 is the only system that exhibits a central object at the expected position of the forming star. The $2.2 \mu\text{m}$ emission is relatively compact ($240 \times 450 \text{ AU}$)—too small to host a cluster. One may speculate whether this elongated feature originates from a binary system that is embedded within a common accretion disk. Assuming that the emission is due to direct stellar light, we derive an absolute infrared brightness of $K \approx -2.5$ mag, which would correspond to a main sequence star of about 20 solar masses. Given the fact that the accretion process is still active, and that models predict that about 30–50% of the circumstellar material can be accumulated onto the central object, it is likely that in the present case a massive protostar is currently being born. Theoretical calculations⁷ show that an initial gas cloud of 60–120 solar masses may evolve into a star between 33 and 43 solar masses, while the remaining mass is rejected into the interstellar medium. □

Methods

Adaptive optics imaging

The sub-arcsec-resolution $2.2 \mu\text{m}$ image was performed with NAOS-CONICA at the ESO VLT in June 2003. Using the background nebular light as a homogeneous source of intensity, the $2.2 \mu\text{m}$ optical depth τ_d can be inferred at each position²³. From τ_d we obtain the projected morphology of the disk. Starting from a central gap of $4 \times 10^{15} \text{ cm}$ radius, this analysis reveals three morphological details: (1) a dense inner torus (delineated by the white contours in Fig. 1) extends up to $r \approx 3.8 \times 10^{16} \text{ cm}$ and widens with increasing radius ($7.9 \times 10^{15} < z < 1.8 \times 10^{16}$), where z is the vertical distance from the disk plane in cm, (2) a flared disk that extends up to $r \approx 9.9 \times 10^{16} \text{ cm}$ ($1.8 \times 10^{16} < z < 4.1 \times 10^{16}$), and (3) an outer envelope of $r \approx 5.3 \times 10^{17} \text{ cm}$ ($1.0 \times 10^{17} < z < 4.0 \times 10^{17}$). The average column density of hydrogen for the disk/envelope system is $\sim 6 \times 10^{22} \text{ atoms cm}^{-2}$ as inferred from $\tau_d \approx 1$.

Millimetre spectroscopy

The molecular line spectroscopy was carried out at the IRAM Plateau de Bure interferometer near Grenoble in April 2003. We have observed the region in the rotational $J = 1 - 0$ transitions of the ^{12}CO , ^{13}CO and C^{18}O molecules, and in the adjacent continuum at 3 mm. Using correlator units of 20 and 40 MHz bandwidth, spectral resolutions of 39 and 78 kHz were achieved, corresponding to velocity resolutions of 0.1 and 0.2 km s^{-1} , respectively. The size of the synthesized beam is about $8'' \times 6''$ with position angles of $40\text{--}60^\circ$. This technique is not sensitive for low spatial frequencies and thus cancels out diffuse foreground/background emission.

Optical and infrared imaging

The images of the disk/hourglass nebula have been obtained with EMMI (0.45, 0.55, $0.8 \mu\text{m}$) at the ESO New Technology Telescope (NTT), La Silla, in July 2003, and with ISAAC (1.25, 1.65 and $2.2 \mu\text{m}$) at the ESO Very Large Telescope (VLT) on Cerro Paranal in

September 2002. The spatial resolution was limited by atmospheric turbulence and varies between $0.4''$ and $0.8''$.

Optical spectroscopy

The optical spectra of the bipolar outflow were measured in April/June 2003 with EFOSC2 at the ESO 3.6 m telescope and with EMMI at the ESO 3.5 m NTT, both located on La Silla, Chile. The aperture lengths were 5' and 8', with a width of 1' and a slit orientation perpendicular to the plane of the disk. The spectra comprise a total wavelength range from 0.4 to $0.9 \mu\text{m}$ and provide a spectral resolution of 2.8 Å per pixel ; the accuracy of the wavelengths calibration allows sampling of velocities down to 40 km s^{-1} .

Received 10 September 2003; accepted 23 March 2004; doi:10.1038/nature02507.

1. Larson, R. The emitted spectrum of a protostar. *Mon. Not. R. Astron. Soc.* **145**, 297–308 (1969).
2. Shu, F., Adams, F. & Lizano, S. Star formation in molecular clouds—Observation and theory. *Annu. Rev. Astron. Astrophys.* **25**, 23–81 (1987).
3. Larson, R. in *Galactic Star Formation Across the Stellar Mass Spectrum* (eds De Buizer, J. M. & van der Blik, N. S.) ASP Conf. Ser. Vol. 287, 65–80 (Astronomical Society of the Pacific, San Francisco, 2003).
4. Wolfire, M. & Cassinelli, J. Conditions for the formation of massive stars. *Astrophys. J.* **319**, 850–867 (1987).
5. Nakano, T., Hasegawa, T. & Norman, C. The mass of the star formed in a cloud core: Theory and its application to the Orion A cloud. *Astrophys. J.* **450**, 183–195 (1995).
6. Jijina, J. & Adams, F. Infall collapse solutions in the inner limit: Radiation pressure and its effects on star formation. *Astrophys. J.* **462**, 874–887 (1996).
7. Yorke, H. & Sonnhalter, C. The formation of massive stars. *Astrophys. J.* **569**, 846–862 (2002).
8. Bonnell, I., Bate, M. & Zinnecker, H. On the formation of massive stars. *Mon. Not. R. Astron. Soc.* **298**, 93–102 (1998).
9. Chini, R. & Wargau, W. Young stellar objects and abnormal extinction within M 17. *Astron. Astrophys.* **329**, 161–168 (1998).
10. Hanson, M. & Conti, P. The young massive stellar objects of M 17. *Astrophys. J.* **489**, 698–718 (1997).
11. Greenhill, L., Gwinn, C., Schwartz, C., Moran, J. & Diamond, P. Coexisting conical bipolar and equatorial outflows from a high-mass protostar. *Nature* **396**, 650–653 (1998).
12. Martí, J., Rodríguez, L. & Reipurth, B. HH 80–81: A highly collimated Herbig-Haro complex powered by a massive young star. *Astrophys. J.* **416**, 208–217 (1993).
13. Hofner, P., Wiesemeyer, H. & Henning, T. A high-velocity molecular outflow from the G9.62 + 0.19 star-forming region. *Astrophys. J.* **549**, 425–432 (2001).
14. Gehr, R. et al. Anatomy of a region of star formation—Infrared images of S106/AFGL 2584. *Astrophys. J.* **254**, 550–561 (1982).
15. Shepherd, D., Claussen, M. & Kurtz, S. Evidence for a solar system-size accretion disk around the massive protostar G192.16–3.82. *Science* **292**, 1513–1518 (2001).
16. Sandell, G., Wright, M. & Forster, J. NGC 7538S—a high-mass protostar with a massive rotating disk. *Astrophys. J.* **590**, L45–L48 (2003).
17. Muzerolle, J., Calvet, N. & Hartmann, L. Emission-line diagnostics of T Tauri magnetospheric accretion. II. Improved model tests and insights into accretion physics. *Astrophys. J.* **550**, 944–961 (2001).
18. Muzerolle, J., Hartmann, L. & Calvet, N. Emission-line diagnostics of T Tauri magnetospheric accretion. I. Line profile observations. *Astron. J.* **116**, 455–468 (1998).
19. Hamann, F. & Persson, S. Emission-line studies of young stars. II. The Herbig Ae/Be stars. *Astrophys. J. Suppl.* **82**, 285–309 (1992).
20. Hartigan, P., Edwards, S. & Ghandour, L. Disk accretion and mass loss from young stars. *Astrophys. J.* **452**, 736–768 (1995).
21. Calvet, N. Properties of the winds of T Tauri stars. *IAU Symp.* **182**, 417–432 (1997).
22. Hartmann, L., Hewett, R. & Calvet, N. Magnetospheric accretion models for T Tauri stars. I. Balmer line profiles without rotation. *Astrophys. J.* **426**, 669–687 (1994).
23. Siebenmorgen, R. & Krügel, E. The protostellar system HH108MMS. *Astron. Astrophys.* **364**, 625–632 (2000).

Acknowledgements We thank the directors of ESO and IRAM for the allocation of discretionary time to perform the adaptive optics image and the molecular line observations, respectively. We also thank the Paranal Science Operations team for performing the infrared observations in service mode. This work was supported by the Nordrhein-Westfälische Akademie der Wissenschaften, funded by the Federal State Nordrhein-Westfalen and the Federal Republic of Germany.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.C. (chini@astro.rub.de).

De Broglie wavelength of a non-local four-photon state

Philip Walther¹, Jian-Wei Pan^{1*}, Markus Aspelmeyer¹, Rupert Ursin¹, Sara Gasparoni¹ & Anton Zeilinger^{1,2}

¹Institut für Experimentalphysik, Universität Wien, Boltzmannngasse 5, 1090 Wien, Austria

²Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Boltzmannngasse 3, 1090 Wien, Austria

* Present address: Physikalisches Institut, Universität Heidelberg, D-69120 Heidelberg, Germany

Superposition is one of the most distinctive features of quantum theory and has been demonstrated in numerous single-particle interference experiments^{1–4}. Quantum entanglement⁵, the coherent superposition of states in multi-particle systems, yields more complex phenomena^{6,7}. One important type of multi-particle experiment uses path-entangled number states, which exhibit pure higher-order interference and the potential for applications in metrology and imaging⁸; these include quantum interferometry and spectroscopy with phase sensitivity at the Heisenberg limit^{9–12}, or quantum lithography beyond the classical diffraction limit¹³. It has been generally understood¹⁴ that in optical implementations of such schemes, lower-order interference effects always decrease the overall performance at higher particle numbers. Such experiments have therefore been limited to two photons^{15–18}. Here we overcome this limitation, demonstrating a four-photon interferometer based on linear optics. We observe interference fringes with a periodicity of one-quarter of the single-photon wavelength, confirming the presence of a four-particle mode-entangled state. We anticipate that this scheme should be extendable to arbitrary photon numbers, holding promise for realizable applications with entanglement-enhanced performance.

To see the origin of multi-particle interference more clearly, consider first a simple analogue to Young's double-slit experiment, that is, a Mach–Zehnder (MZ) interferometer (Fig. 1a). There, single-photon interference occurs owing to the spatial separation of two modes of propagation a1 and b1 for a single particle entering the interferometer at the first beamsplitter. Variation of the path length induces a phase shift $\Delta\varphi$ and thus gives rise to detection probabilities $P_{a2} \propto 1 + \cos\Delta\varphi$ and $P_{b2} \propto 1 - \cos\Delta\varphi$ in each of the two output modes a2 and b2 behind the exit beamsplitter. Two-photon interference occurs when a1 and b1 are the modes of propagation for a state of two indistinguishable photons, that is, a biphoton state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|2\rangle_{a1}|0\rangle_{b1} + e^{i2\Delta\varphi}|0\rangle_{a1}|2\rangle_{b1})$. It represents a path-entangled two-photon state, which exhibits pure two-particle interference at the output beamsplitter. The probability to find two photons in either mode a2 or b2 then oscillates with $P_{a2,a2} \propto 1 + \cos(2\Delta\varphi)$ and $P_{b2,b2} \propto 1 - \cos(2\Delta\varphi)$, respectively, while the single-photon detection probabilities P_{a2} and P_{b2} remain constant.

In the generalized case of an N -particle interferometer, the N photons will be in a superposition of being in either mode a1 or b1, resulting in

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|N\rangle_{a1}|0\rangle_{b1} + e^{iN\Delta\varphi}|0\rangle_{a1}|N\rangle_{b1}) \quad (1)$$

In other words, the paths are entangled in photon number. Here $|N\rangle_{a1}(|N\rangle_{b1})$ indicates the N -particle Fock state in spatial mode a1 (b1), respectively, and $N = 0$ represents an empty mode. The phase modulation $N\Delta\varphi$ increases linearly with the particle number N , which is the origin of all entanglement-enhanced interferometric schemes. In particular, the N -photon detection probability in each of the interferometer outputs would vary as $P_N \propto 1 \pm \cos(N\Delta\varphi)$. It

has therefore been suggested¹⁹ that an effective de Broglie wavelength λ/N could be attributed to the quantum state. This resembles the case of a heavy massive molecule⁴ consisting of N atoms; though here the particles are in no way bound to each other.

In order to benefit best from such entanglement-enhanced interferometric techniques, it is desirable to achieve experimentally a high photon number N for states of the form of equation (1). The special case of $N = 2$ was realized both in the original Young's double-slit geometry (by using collinear production of biphoton states via parametric down conversion²⁰), and in the Mach–Zehnder configuration (by using two-photon interference to suppress unwanted single-photon contributions^{15–18}). It is commonly believed that the realization for states with $N > 2$ requires the use of nonlinear gates²¹ or N additional 'ancilla' detectors with single-photon resolution²². Unfortunately, neither of these schemes is feasible with current technologies. We demonstrate how to overcome this limitation, giving a specific example of pure four-photon interferometry.

Our proposal is based on separating photon pairs into different pairs of modes, and using two-particle interferometry rather than distinguishing photon numbers or using nonlinear beamsplitters. To achieve this goal, we exploit type-II spontaneous parametric down-conversion (SPDC)²³. An ultraviolet (UV) pulse passes through a β -barium-borate crystal, probabilistically emitting pairs of energy-degenerate polarization-entangled photons into the spatial modes a1 and a2 (Fig. 1b). The UV pump beam is reflected back at a mirror, and can thus also emit entangled photon pairs into the spatial modes b1 and b2. The set-up is aligned to generate the following maximally entangled biphoton state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|H\rangle_{a1}|H\rangle_{b2} + |V\rangle_{a1}|V\rangle_{b2}) \quad (2)$$

for each of the pairs emitted into the pairs of modes a1–a2 or b1–b2. Here H (V) indicates horizontal (vertical) polarization of the photon.

We first consider the case where only one pair of entangled photons is emitted on a double pass of the UV pulse through the crystal. There are two probability amplitudes that will contribute to the emerging two-photon state; the pair is emitted either into the pair of modes a1–a2 or into the pair of modes b1–b2. We then coherently combine the two pairs of modes at the two polarizing beamsplitters (PBSs). As the PBS transmits horizontally polarized light and reflects vertically polarized light, conditional on detecting one photon in each of the output ports, for example, a3 and a4, the biphoton state will be

$$|\Phi\rangle_{a3a4} = \frac{1}{\sqrt{2}}(|H\rangle_{a3}|H\rangle_{a4} + e^{i2\Delta\varphi}|V\rangle_{a3}|V\rangle_{a4}) \quad (3)$$

where again $\Delta\varphi$ is the phase modulation of a single photon^{15,24}. The phase $\Delta\varphi$ can be modulated by changing the position of the pump mirror. Two-photon interference fringes may now be observed by performing a projection measurement in the modes a3 and a4 into the linear polarization basis $|\pm\rangle = (1/\sqrt{2})(|H\rangle \pm |V\rangle)$. Specifically, the probability of detecting a twofold coincidence $|+\rangle_{a3}|+\rangle_{a4}$ is proportional to $P_{a3,a4} \propto 1 - \cos(2\Delta\varphi)$. These correlations are already a signature of non-locality^{25,26} of the two-photon state (Fig. 2b).

We now explain how our scheme can be generalized to four photons and even higher photon numbers. Consider the case where four photons are emitted on a double pass of the pump beam through the crystal. There are two possibilities that will contribute to an overall four-photon state: first, a double-pair is emitted on either pass of the pump beam, that is, two photon pairs are emitted into a superposition of being in the same pair of modes a1–a2 or b1–b2, respectively, or, second, one pair of photons is emitted on each pass of the pump beam, that is, into each of the modes a1–a2 and b1–b2.

We first study the double-pair emission case. In our experiment, the spectral filtering in the photon detection is much narrower than the linewidth of the UV-pump, so the two pairs cannot be treated as independent but have to be described as one four-photon²⁷. In other words, the double-pair emission results in a four-photon propagating within the mode pair a1–a2 or b1–b2. A fourfold coincidence after the two PBS, that is, detection of a single photon in each of the output ports a3, a4, b3 and b4, will either result from a $|H\rangle_{a3}|H\rangle_{a4}|V\rangle_{b3}|V\rangle_{b4}$ contribution, if the double pair is in a1–a2, or from a $|V\rangle_{a3}|V\rangle_{a4}|H\rangle_{b3}|H\rangle_{b4}$ contribution, if the double pair is in b1–b2. Temporal overlapping of both pairs of modes at the two

PBS results in $|H\rangle_{a3}|H\rangle_{a4}|V\rangle_{b3}|V\rangle_{b4} + e^{i4\Delta\varphi}|V\rangle_{a3}|V\rangle_{a4}|H\rangle_{b3}|H\rangle_{b4}$, a coherent superposition of forward and backward emission at the same time, where all the four backward emitted photons are phase shifted by the pump mirror. Introducing a path difference induces a phase shift $\Delta\varphi$, and further performing a polarization measurement in the $|\pm\rangle$ basis to achieve indistinguishability results in interference fringes with one-quarter of the single-photon wavelength. Note again that it originates from an intrinsic four-photon effect and does not involve lower-order interference. We will now show how this four-photon effect was isolated in the experiment without unwanted lower-order interference.

With the same probability as the desired double-pair emission, one photon pair is emitted into each of the mode pairs a1–a2 and b1–b2 by one pulse. The resulting coherent superposition $e^{i2\Delta\varphi}(|H\rangle_{a3}|H\rangle_{a4}|H\rangle_{b3}|H\rangle_{b4} + |V\rangle_{a3}|V\rangle_{a4}|V\rangle_{b3}|V\rangle_{b4})$ has a fixed relative phase²⁸ which is independent of the position of the pump mirror. Also in contrast to the above four-photon emission, only two-photon interference contributes to these events. These contributions can be erased by performing a proper projection measurement of the four output modes a3, b3, a4 and b4 into the $|\pm\rangle$ bases; then the number of $|+\rangle$ projections is different from the number of $|-\rangle$ projections, say $|+\rangle_{a3}|-\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4}$. The overall four-photon amplitude originating from one photon in each mode then becomes $e^{i2\Delta\varphi}(|+\rangle_{a3}|-\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4} - |+\rangle_{a3}|-\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4})$ and thus vanishes owing to destructive interference, a four-photon equivalent to a Hong-Ou-Mandel interference²⁹.

Figure 2 compares the resulting pure four-photon interference effect (Fig. 2c) with the well-known two-photon interference (Fig. 2b) and the single-photon interference (Fig. 2a) obtained with the same set-up. The data reveal a reduction of the oscillation wavelength, from 823 ± 46 nm for the single-photon case, via 395 ± 16 nm for the two-photon case, to 194 ± 9 nm for the four-photon case. The result demonstrates that one has to treat the four-photon state as one object of the form $|\psi\rangle = \frac{1}{\sqrt{2}}(|4\rangle_{a1,a2}|0\rangle_{b1,b2} + e^{i4\Delta\varphi}|0\rangle_{a1,a2}|4\rangle_{b1,b2})$, which is similar to a 'NOON'-state³⁰. Our four-photon state has the additional interesting property that it is non-local, as it is a superposition of four photons either in mode a1 and a2 or b1 and b2.

Nevertheless there is room for an ambiguous interpretation of the observed four-photon oscillations via two-photon interference

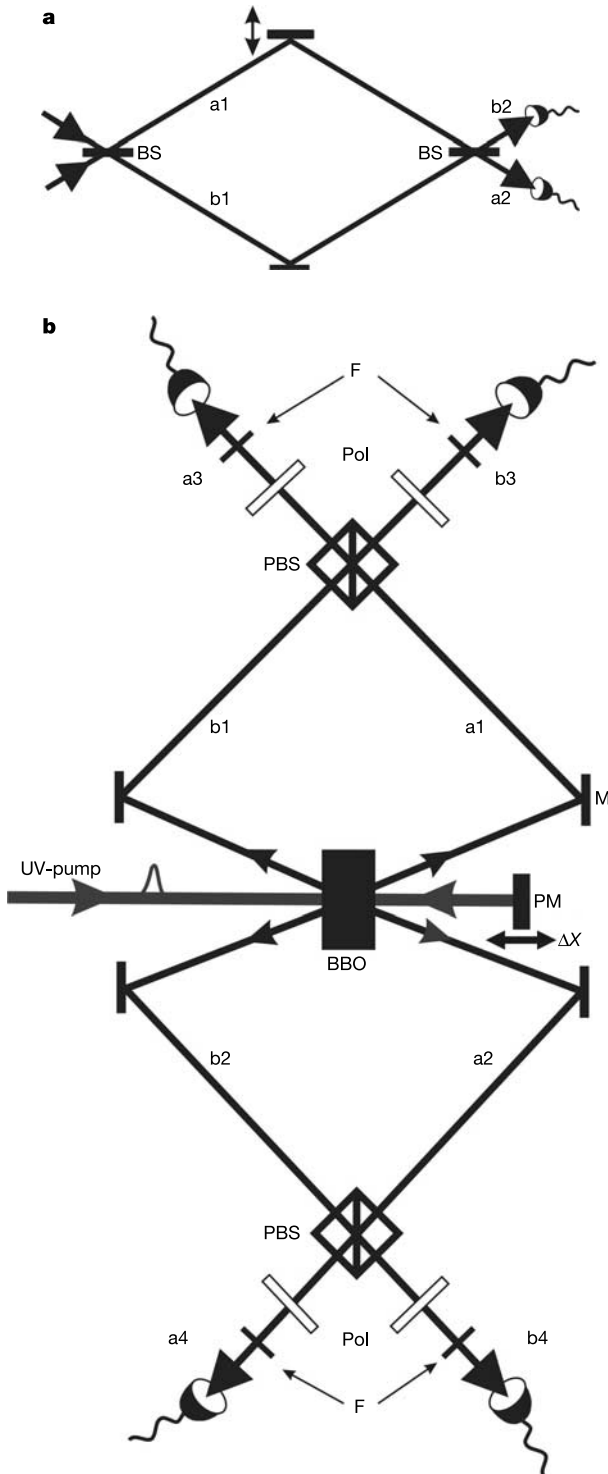


Figure 1 One-, two- and four-photon interferometry. **a**, In a two-mode Mach-Zehnder interferometer, the phase is changed by varying the path length via the position of a mirror. Single-photon interference occurs owing to the spatial separation of two possible modes of propagation a1 and b1 for a single particle entering the interferometer at the first beamsplitter (BS). Two-photon interference can be achieved when a1 and b1 are the two possible modes of propagation for a biphoton state. **b**, In our experiment, the required four-photon state is produced by type-II spontaneous parametric down-conversion (SPDC). A 200 fs pulse at a central UV wavelength of 395 nm and at a repetition rate of 76 MHz passes through a β -barium-borate (BBO) crystal, probabilistically emitting pairs of energy-degenerate polarization-entangled photons at 790 nm into the spatial modes a1 and a2. The UV pump beam is reflected back at a mirror, and might thus emit a second pair into the spatial modes b1 and b2. The probability of single-pair creation is of the order of p (in our set-up, $p \approx 10^{-2}$ – 10^{-3}), while the probability of creating two pairs is proportional to p^2 . Filters (F) of 3 nm bandwidth coupling into single-mode fibres in front of each detector enable good temporal and spatial overlap of the photon wavepackets at the polarizing beamsplitters (PBSs). The UV pump is reflected by the pump mirror PM, which is mounted on a computer-controlled translation stage. By scanning the position of PM with a step size of 1 μ m and performing fine adjustment of the position of M, we achieved the temporal overlap of modes a1 and b1, and of modes a2 and b2. An additional piezo translation stage is used to move the pump mirror PM and to perform a change of the phase between four photons emitted into modes a1 and a2 relative to the four photons emitted into b1 and b2. The detection of the spatially separated four-photon coincidences behind a 45° polarizer (Pol) while varying the position of PM leads to the observed interference fringes.

effects. However, this loophole can be definitely closed by the following experiment, in which we observe the desired four-photon interference without any signature of lower-order interference effects. If the assumption that the fourfold coincidence pattern of Fig. 2 could be reduced to two-photon interference were correct,

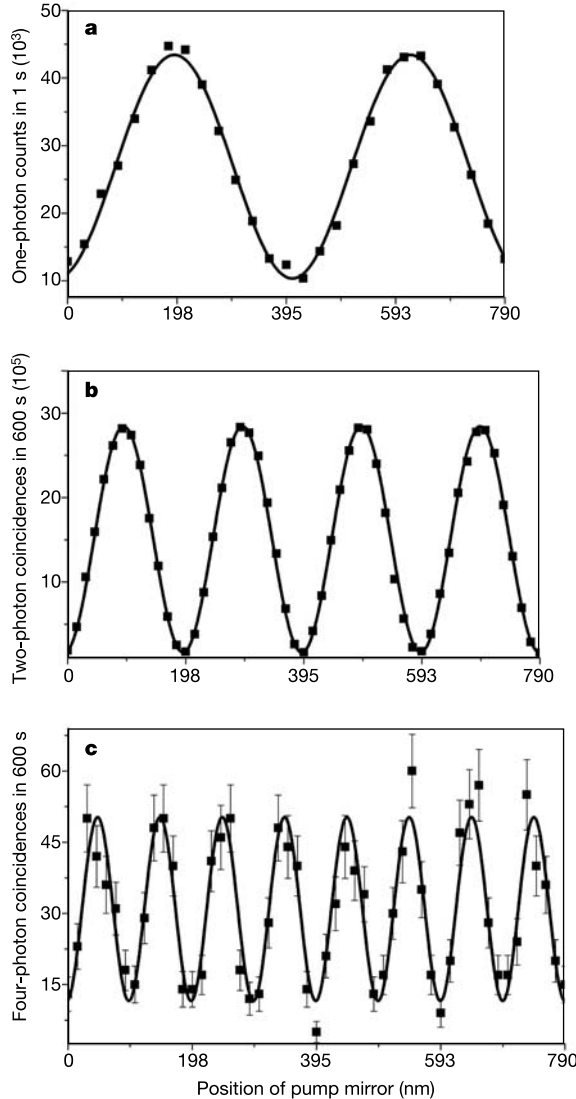


Figure 2 Experimental demonstration of pure one-, two- and four-photon interference. The two- and four-photon interference is recorded simultaneously, whereas for the one-photon interferometry the pulsed laser has been switched from mode-locking to continuous-wave (c.w.) mode. **a**, The single photon rate in mode a3 after performing a projection measurement in the linear polarization basis $|\pm\rangle = (1/\sqrt{2})(|H\rangle \pm |V\rangle)$. For this interference pattern, the pump laser is used in the c.w. mode at 790 nm (instead of the mode-locked frequency-doubled mode at 395 nm). A Mach-Zehnder configuration for modes a1–b1 arises for light scattered from the BBO crystal when passing through the crystal. By moving the pump mirror, interference fringes appear for single photons with a central wavelength of 790 nm, which corresponds to the down-converted photons. Note that, owing to the back reflection of the pump beam, the change in the optical path is twice as large as in the position of the pump mirror. **b**, The two-photon coincidence rate corresponding to the detection in mode a3 and a4 after projecting onto $|+\rangle_{a3}|-\rangle_{a4}$. **c**, A demonstration of how performing a projection onto $|+\rangle_{a3}|-\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4}$ results in pure four-photon interference owing to projection onto the (non-local) path-entangled four-photon state $|\psi\rangle = \frac{1}{\sqrt{2}}(|\psi\rangle_{a1,a2}|\psi\rangle_{b1,b2} + e^{i4\Delta\varphi}|\psi\rangle_{a1,a2}|\psi\rangle_{b1,b2})$. The visibility of the observed four-photon oscillations, defined as $V = (I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$, is approximately 61%. The error bars in Figs 2 and 3 are defined as the square root of the observed fourfold coincidence.

then a suppression of the two-photon amplitudes should also lead to a suppression of the fourfold oscillation. The experimental data clearly rules out this scenario (Fig. 3). In this experiment, the initial states are prepared such that a forward emitted pair is in state $|\Phi^+\rangle$ as before, while a backward emitted pair is in the state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|H\rangle_{b1}|V\rangle_{b2} - |V\rangle_{b1}|H\rangle_{b2})$. A superposition of these states can never result in any two-photon interference owing to the bilateral parity check²⁴ performed by the two PBSs: for the $|\Phi^+\rangle$ -Bell state both photons are always transmitted or reflected, whereas for the $|\Psi^-\rangle$ -Bell state one photon is always transmitted and one is reflected, such that no interfering amplitudes for twofold detection events can build up. Only a double-pair emission on each side, where a four-photon is emitted either forwards or backwards, contributes to the four-photon state after the PBS and gives rise to pure four-photon interference. In both cases of a $|+\rangle_{a3}|-\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4}$ -projection or a $|+\rangle_{a3}|+\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4}$ -projection, all possible two-photon coincidence detections do not show any oscillations (Fig. 3a), while the four-photon coincidence rate oscillates with $P_{a3,a4,b3,b4} \propto 1 - \cos(4\Delta\varphi)$ or $P_{a3,a4,b3,b4} \propto 1 + \cos(4\Delta\varphi)$, respectively, that is, a fourfold reduction in wavelength (Fig. 3b). This complete absence of two-photon interference effects underlines the true four-photon character of the observed interference. Note that this is in contrast to the first case, in which a projection onto $|+\rangle_{a3}|+\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4}$ would not eliminate the two-photon interference and thus does not

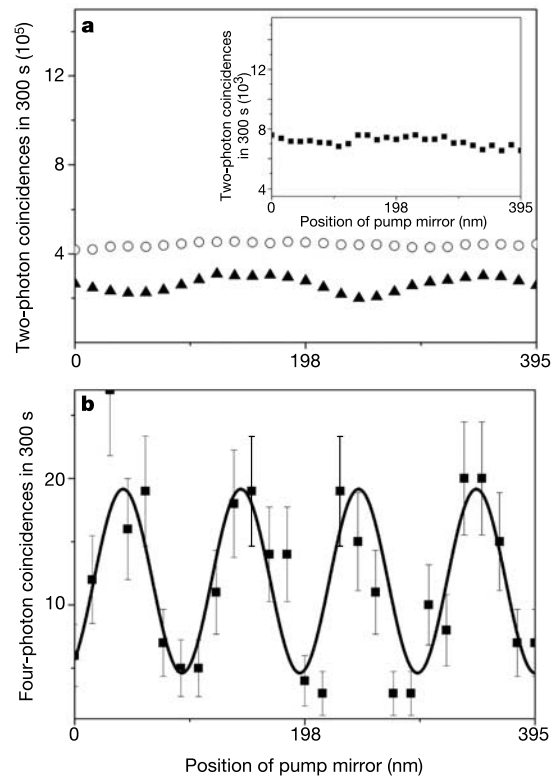


Figure 3 Pure four-photon interference without any two-photon interference. This can be observed when different entangled input states are used in forward and backward emission, specifically $|\Phi^+\rangle_{a1,a2}$ and $|\Psi^-\rangle_{b1,b2}$. Owing to the bilateral parity action of the PBS, only a four-photon emitted in a superposition of either forward or backward direction can contribute to the overall state. **a**, No interference is obtained from any possible twofold coincidence measurement $P_{a3,a4}$ (open circles), $P_{a4,b3}$ (filled triangles) and $P_{a4,b4}$ (inset). **b**, In the case of a $|+\rangle_{a3}|+\rangle_{a4}|+\rangle_{b3}|+\rangle_{b4}$ -projection, pure four-photon oscillations can be observed with detection probability $P_{a3,a4,b3,b4} \propto 1 - \cos(4\Delta\varphi)$. The corresponding measured visibilities of the two-photon curves are approximately 4% for $P_{a3,a4}$, 19% for $P_{a4,b3}$ and 7% for $P_{a4,b4}$, which is clearly below the four-photon fringe visibility of approximately 61%.

result in the desired $\lambda/4$ de Broglie wave effect³¹. In principle, this can be extended to higher-order interference effects because, obviously, when more than two double-pairs are emitted from the crystal the suppression of all lower-order interferences can be achieved by a proper projection analogous to the $N = 4$ case.

The method that we use here allows the generation of four-photon states and their subsequent utilization in pure four-particle interferometry. The result clearly confirms the theoretical expectation that the de Broglie wavelength of a four-photon state is one-quarter that of a single photon, thus leading to the general rule $\lambda(N) = \lambda(1)/N$. This overcomes the resolution limit of state-of-the-art two-particle interferometry, and opens new possibilities for multi-particle interference in fundamental quantum experiments and in applications such as quantum metrology. It is important to note that, in principle, this scheme can be extended to higher particle numbers if more spatial modes are involved. The actual limitation due to low count rates may eventually be overcome with the next generation of entangled photon sources and detectors. □

Received 14 November 2003; accepted 6 April 2004; doi:10.1038/nature02552.

1. Marton, L., Simpson, J. A. & Suddeth, J. A. Electron beam interferometer. *Phys. Rev.* **90**, 490–491 (1953).
2. Rauch, H., Treimer, W. & Bonse, U. Test of a single crystal neutron interferometer. *Phys. Lett. A* **47**, 369–371 (1974).
3. Keith, D. W., Ekstrom, C. R. & Pritchard, D. E. An interferometer for atoms. *Phys. Rev. Lett.* **66**, 2693–2696 (1991).
4. Arndt, M. *et al.* Wave-particle duality of C_{60} molecules. *Nature* **401**, 680–682 (1999).
5. Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**, 807–812, 823–828, 844–849 (1935).
6. Horne, M. A. & Zeilinger, A. in *Symposium on the Foundations of Modern Physics* (eds Joensuu, L. P. & Mittelstedt, P.) 435–439 (World Scientific Press, Singapore, 1985).
7. Greenberger, D., Horne, M. & Zeilinger, A. Multiparticle interferometry and the superposition principle. *Phys. Today* **8**, 22–29 (1993).
8. Lee, H., Kok, P. & Dowling, J. P. in *Proc. Sixth Int. Conf. on Quantum Communication, Measurement and Computing* (eds Shapiro, J. H. & Hirota, O.) 223–229 (Rinton Press, Princeton, 2002).
9. Yurke, B. Input states for enhancement of fermion interferometer sensitivity. *Phys. Rev. Lett.* **56**, 1515–1517 (1986).
10. Ou, Z. Y. Fundamental quantum limit in precision phase measurement. *Phys. Rev. A* **55**, 2598–2609 (1997).
11. Holland, M. J. & Burnett, K. Interferometric detection of optical phase shifts at the Heisenberg limit. *Phys. Rev. Lett.* **71**, 1355–1358 (1993).
12. Bollinger, J. J., Itano, W. M., Wineland, D. J. & Heinzen, D. J. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A* **54**, R4649–R4652 (1996).
13. Boto, A. *et al.* Quantum interferometric optical lithography: Exploiting entanglement to beat the diffraction limit. *Phys. Rev. Lett.* **85**, 2733–2736 (2000).
14. de Steuermagel, O. Broglie wavelength reduction for a multiphoton wave packet. *Phys. Rev. A* **65**, 033820 (2002).
15. Ou, Z. Y., Wang, L. J., Zou, X. Y. & Mandel, L. Evidence for phase memory in two-photon down conversion through entanglement with the vacuum. *Phys. Rev. A* **41**, 566–568 (1990).
16. Rarity, J. G. Two-photon interference in a Mach-Zehnder interferometer. *Phys. Rev. Lett.* **65**, 1348–1351 (1990).
17. Ou, Z. Y., Zou, X. Y., Wang, L. J. & Mandel, L. Experiment on nonclassical fourth-order interference. *Phys. Rev. A* **42**, 2957–2965 (1990).
18. Edamatsu, K., Shimizu, R. & Itoh, T. Measurement of the photonic de Broglie wavelength of entangled photon pairs generated by spontaneous parametric down-conversion. *Phys. Rev. Lett.* **89**, 213601 (2002).
19. Jacobson, J., Björk, G., Chuang, I. & Yamamoto, Y. Photonic de Broglie waves. *Phys. Rev. Lett.* **74**, 4835–4838 (2002).
20. Fonseca, E. J. S., Monken, C. H. & Padua, S. Measurement of the de Broglie wavelength of a multiphoton wave packet. *Phys. Rev. Lett.* **82**, 2868–2871 (1999).
21. Gerry, C. C. & Campos, R. A. Generation of maximally entangled photonic states with a quantum-optical Fredkin gate. *Phys. Rev. A* **64**, 063814 (2001).
22. Kok, P., Lee, H. & Dowling, J. Creation of large-photon number path entanglement conditioned on photodetection. *Phys. Rev. A* **65**, 052104 (2002).
23. Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
24. Pan, J.-W., Gasparoni, S., Ursin, R., Weihs, G. & Zeilinger, A. Experimental entanglement purification of arbitrary unknown states. *Nature* **423**, 417–422 (2003).
25. Ou, Z. Y. & Mandel, L. Violation of Bell's inequality and classical probability in a two-photon correlation experiment. *Phys. Rev. Lett.* **61**, 50–53 (1988).
26. Shih, Y. H. & Alley, C. O. New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion. *Phys. Rev. Lett.* **61**, 2921–2924 (1988).
27. Weinfurter, H. & Zukowski, M. Four-photon entanglement from down-conversion. *Phys. Rev. A* **64**, 010102 (2001).
28. Simon, C. & Pan, J.-W. Polarization entanglement purification using spatial entanglement. *Phys. Rev. Lett.* **89**, 257901 (2002).
29. Hong, C. K., Ou, Z. Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
30. Kok, P. *et al.* Quantum interferometric optical lithography: Towards arbitrary two-dimensional patterns. *Phys. Rev. A* **63**, 063407 (2001).

31. Lamas-Linares, A., Howell, J. C. & Bouwmeester, D. Stimulated emission of polarization-entangled photons. *Nature* **412**, 887–890 (2001).

Acknowledgements We thank Č. Brukner and K. Resch for discussions, and V. Scarani for comments on the manuscript. This work was supported by the Austrian Science Foundation (FWF), the European Commission in the RAMBOQ project and by the Alexander von Humboldt Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.Z. (zeilinger-office@quantum.at).

Super-resolving phase measurements with a multiphoton entangled state

M. W. Mitchell, J. S. Lundeen & A. M. Steinberg

Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario M5S 1A7, Canada

Interference phenomena are ubiquitous in physics, often forming the basis of demanding measurements. Examples include Ramsey interferometry in atomic spectroscopy, X-ray diffraction in crystallography and optical interferometry in gravitational-wave studies^{1,2}. It has been known for some time that the quantum property of entanglement can be exploited to perform super-sensitive measurements, for example in optical interferometry or atomic spectroscopy^{3–7}. The idea has been demonstrated for an entangled state of two photons⁸, but for larger numbers of particles it is difficult to create the necessary multiparticle entangled states^{9–11}. Here we demonstrate experimentally a technique for producing a maximally entangled three-photon state from initially non-entangled photons. The method can in principle be applied to generate states of arbitrary photon number, giving arbitrarily large improvement in measurement resolution^{12–15}. The method of state construction requires non-unitary operations, which we perform using post-selected linear-optics techniques similar to those used for linear-optics quantum computing^{16–20}.

Our goal is to create the state

$$|N::0\rangle_{a,b} \equiv \frac{1}{\sqrt{2}}(|N,0\rangle_{a,b} + |0,N\rangle_{a,b}) \quad (1)$$

which describes two modes a, b in a superposition of distinct Fock states $|n_a = N, n_b = 0\rangle$ and $|n_a = 0, n_b = N\rangle$. This state figures in several metrology proposals, including atomic frequency measurements⁴, interferometry^{3,6,7}, and matter-wave gyroscopes⁵. In these proposals the particles occupying the modes are atoms or photons.

The advantage for spectroscopy can be seen in this idealization: we wish to measure a level splitting $H_{\text{ext}} = \epsilon_{ba}b^\dagger b$ between modes b and a using a fixed number of particles N in a fixed time T . We could prepare N copies of the single-particle state $(|1,0\rangle_{a,b} + |0,1\rangle_{a,b})/\sqrt{2}$ and allow them to evolve to the state $|\phi\rangle \equiv (|1,0\rangle_{a,b} + \exp[i\phi]|0,1\rangle_{a,b})/\sqrt{2}$, where $\phi = \epsilon_{ba}T/\hbar$. Measurements of $A_1 \equiv |0,1\rangle\langle 1,0| + |1,0\rangle\langle 0,1|$ on this ensemble give $\langle A_1 \rangle = \cos(\phi)$ with phase uncertainty at the shot-noise limit, $\Delta\phi = 1/\sqrt{N}$. In contrast, under the same hamiltonian $|N::0\rangle$ evolves to $(|N,0\rangle + \exp[iN\phi]|0,N\rangle)/\sqrt{2}$. If we measure the operator $A_N \equiv$

$|0, N\rangle \langle N, 0| + |N, 0\rangle \langle 0, N|$, we find $\langle A_N \rangle = \cos(N\phi)$. The dependence on $N\phi$ rather than ϕ is phase super-resolution: one cycle of $\langle A_N \rangle$ implies a smaller change of ϕ (or ϵ_{ba}) than one cycle of $\langle A_1 \rangle$. Phase super-sensitivity, a reduction of phase uncertainty, is also predicted. A number of schemes have been proposed^{3–7,12,21,22} to reach the so-called Heisenberg limit $\Delta\phi = 1/N$. The simplest proposals would measure the operator A_N . This can be implemented with coincidence measurements, as the probability of detecting all N quanta in a mode $(a+b)/\sqrt{2}$ is proportional to $1 + \langle A_N \rangle$. Phase super-resolution has been demonstrated for $N=2$ with photons in a Mach–Zehnder interferometer^{23,24} and with trapped ions²⁵.

A related technique, quantum interferometric optical lithography, proposes using phase super-resolution to write features smaller than the single-photon diffraction limit. There the modes a , b are spatial, with different propagation directions. A molecule exposed to both modes and capable of absorbing N photons would, in effect, perform the measurement of A_N as above, with N -fold super-resolution in position. Using coincidence detection in place of two-photon absorbers, this principle has been demonstrated for $N=2$ using down-converted pairs⁸. In that experiment, as well as in the earlier Mach–Zehnder experiments, two infrared photons showed the same resolution or angular resolution as the blue pump photon that generated them, a factor of two improvement over the resolution of a single infrared photon. The question remains as to whether resolution can be improved beyond that of the photons used to generate the entangled state. Here we answer that question in the affirmative by constructing a multiparticle state with greater phase resolution than any of its single-particle precursors. The technique could in principle be used to generate entangled states of arbitrarily large N with arbitrarily good resolution.

We prepare the state $|3::0\rangle_{a,b}$ where the modes a and b are the horizontal (H) and vertical (V) polarizations of a single spatial mode. The construction of the polarization state is based on earlier proposals to construct photon-number path-entangled states^{12–14}. A similar technique for polarization has recently been independently proposed¹⁵. The key to the construction is the fact that $|3::0\rangle_{a,b}$, when written in terms of creation operators $a_a^\dagger$ and $a_b^\dagger$ acting on the vacuum $|0\rangle$, is equal to

$$|3::0\rangle_{a,b} = (a_a^\dagger + a_b^\dagger)(a_a^\dagger + e^{i\chi}a_b^\dagger)(a_a^\dagger + e^{i2\chi}a_b^\dagger)|0\rangle \quad (2)$$

where $\chi = 2\pi/3$ and normalization has been omitted. The terms in parentheses each create a particle, but in non-orthogonal states. If a and b are left (L) and right (R) circular polarization, these states describe linear polarizations rotated by 60° from each other. Using post-selection, we can put one photon of each polarization into a single spatial mode and create $|3::0\rangle_{a,b}$.

We use two photons from pulsed parametric down-conversion plus one laser, or ‘local oscillator’ (LO), photon, adapting a technique first used to show non-classical interference of independent sources^{26,27}. Pulses of 100 fs duration and 810 nm centre wavelength are produced by a mode-locked Ti:sapphire laser and frequency-doubled, giving 405-nm pulses with an average power of 50 mW. These are gently focused into a 0.5-mm-thick β -barium borate crystal aligned for type-II parametric down-conversion in the ‘collapsed cone’ geometry²⁸. The down-converted (DC) photons thus produced are orthogonally polarized. A small part of the Ti:sapphire beam is split off and attenuated to contribute the LO photon. These three photons are transformed into the state $|3::0\rangle$ and detected by polarization-resolved threefold coincidence detection. The transformation (Fig. 1a) can be understood as a sequence of mode combinations. After the PBS, the DC photons are in the state $a_{+45^\circ}^\dagger a_{-45^\circ}^\dagger |0\rangle$, where subscripts indicate polarization. A half wave-plate rotates this to $a_{+45^\circ}^\dagger a_{-45^\circ}^\dagger |0\rangle$, where subscripts show linear polarizations measured from vertical. We perform the first post-selected non-unitary operation by passing the pair through three

plates of BK7 glass near the Brewster angle. The six Brewster-angle interfaces act as a partial polarizer (PP) with transmission efficiencies of $T_H \approx 1$, $T_V \approx 1/3$. By post-selecting cases where no photons are reflected, we transform the state as (again without normalization):

$$\begin{aligned} a_{+45^\circ}^\dagger a_{-45^\circ}^\dagger |0\rangle &= (a_H^\dagger + a_V^\dagger)(a_H^\dagger - a_V^\dagger)|0\rangle \\ &\rightarrow (a_H^\dagger + \frac{1}{\sqrt{3}}a_V^\dagger)(a_H^\dagger - \frac{1}{\sqrt{3}}a_V^\dagger)|0\rangle \quad (3) \\ &= a_{+60^\circ}^\dagger a_{-60^\circ}^\dagger |0\rangle \end{aligned}$$

This operation, putting orthogonally polarized photons into non-orthogonal modes, is non-unitary and requires post-selection.

The DC photons meet the LO photon at the last interface of the PP. This interface acts as a beamsplitter putting all three into the same spatial mode, conditioned on zero photons leaving by the ‘dark’ port. Again, the operation is non-unitary and requires post-selection. The LO photon is vertically polarized, and the state thus constructed is $a_{+60^\circ}^\dagger a_{-60^\circ}^\dagger a_{-60^\circ}^\dagger |0\rangle$. This state has sixfold rotational symmetry about the beam propagation direction, and thus can only contain the states $|3, 0\rangle_{L,R}$ and $|0, 3\rangle_{L,R}$, where subscripts indicate the circular polarization basis¹⁵. In fact, it is easily verified that the state is $(|3, 0\rangle_{L,R} - |0, 3\rangle_{L,R})/\sqrt{2}$ up to an overall phase. Finally, we convert circular to linear polarizations with a quarter wave-plate, giving $(|3, 0\rangle_{H,V} - |0, 3\rangle_{H,V})/\sqrt{2}$. Ideally, the probability of both post-selections succeeding is $2 \cos^4(\pi/12)/3 \approx 58\%$.

Response to the phase shifter demonstrates phase super-resolution. Acting on a single photon, the quartz wedges shift by ϕ the phase of the V-polarization. Acting on three photons the phase shift is tripled: $|3::0\rangle_{H,V}$ becomes $|3::0\rangle_{H,V}^{3\phi} = (|3, 0\rangle_{H,V} + \exp[i3\phi]|0, 3\rangle_{H,V})/\sqrt{2}$, where we have absorbed the negative sign above into the phase factor. The 3ϕ behaviour can be seen in threefold coincidence detection (triples detection) in the $\pm 45^\circ$

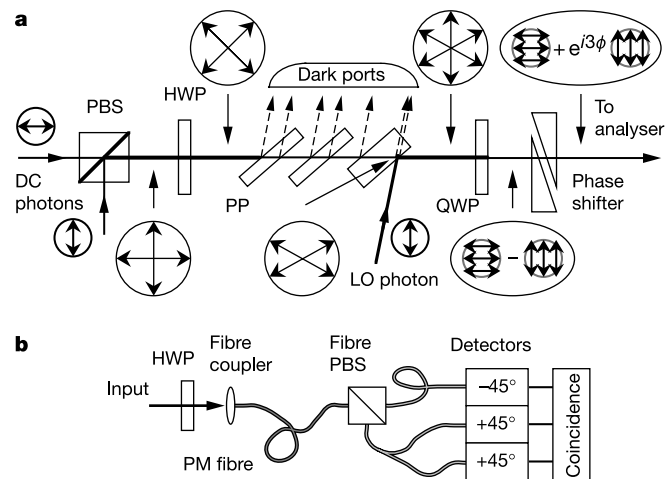


Figure 1 Diagram of production and detection of the state $|3::0\rangle_{H,V}^{3\phi}$. **a**, The chain of optical components and their effect on polarization state, represented in encircled figures. All photons have wavelength 810 nm. A polarizing beamsplitter (PBS) puts the DC photons in a single spatial mode and a half wave-plate (HWP) rotates their polarizations to $\pm 45^\circ$. A partial polarizer (PP) transforms the polarizations to $\pm 60^\circ$ if no photons are reflected into the dark ports. The LO photon is injected at the final interface of the partial polarizer. A quarter wave-plate (QWP) rotates to $(|3, 0\rangle_{H,V} - |0, 3\rangle_{H,V})/\sqrt{2}$ and quartz wedges produce an adjustable phase 3ϕ between the two components of the state. **b**, Analysis in the $\pm 45^\circ$ polarization basis is performed with a HWP before a polarization-maintaining (PM) fibre and a fibre-coupled PBS. The outputs of the fibre PBS are channelled to one, two or three detectors as needed. The configuration to detect $|2, 1\rangle_{\pm 45^\circ}$ is shown. Digital electronics record single detections as well as two- and three-fold coincidences.

linear polarization basis. The rates for detection of $|3, 0\rangle_{\pm 45^\circ}$ and $|2, 1\rangle_{\pm 45^\circ}$ vary as $1 \pm \cos(3\phi)$, respectively. After passing through an 810 nm wavelength filter with a 10 nm passband, the photons enter the polarization analyser, set to detect $|3, 0\rangle_{\pm 45^\circ}$ or $|2, 1\rangle_{\pm 45^\circ}$ (Fig. 1b).

The use of DC pairs removes the need for detectors at the 'dark' ports. Down-conversion very infrequently produces more than two photons in a single pulse, so we can infer with near-certainty the absence of photons in the 'dark' port from the presence of both photons in the 'bright' port. Using a weak coherent state to supply the third photon, we can make small the probability that more than one LO photon was present in a triple detection event¹⁴. Thus a single post-selection for threefold coincidence at the detectors performs at once the post-selections for both non-unitary operations.

Figure 2 shows results for detection of multiple polarizations at the analyser. Intensities of the DC and LO sources were adjusted such that one-photon detections (singles detections) are mostly produced by LO photons (about 10:1 ratio versus DC singles), twofold coincidences mostly by DC pairs (about 5:1 versus LO accidentals), and threefold coincidences principally by one LO photon plus one DC pair (about 2:1 versus accidental triples contributions, below). Thus with a single scan of the phase shifter we can see qualitatively different behaviours for states of one, two and three photons. Figure 2c clearly shows oscillation with 3ϕ , as predicted by theory. The resolution exceeds that achievable with any single photon in the experiment. The 405-nm photons would show oscillation with 2.1ϕ owing to their shorter wavelength and the somewhat larger birefringence of quartz at that wavelength. The observed 3ϕ oscillation is analogous to multipath interference:

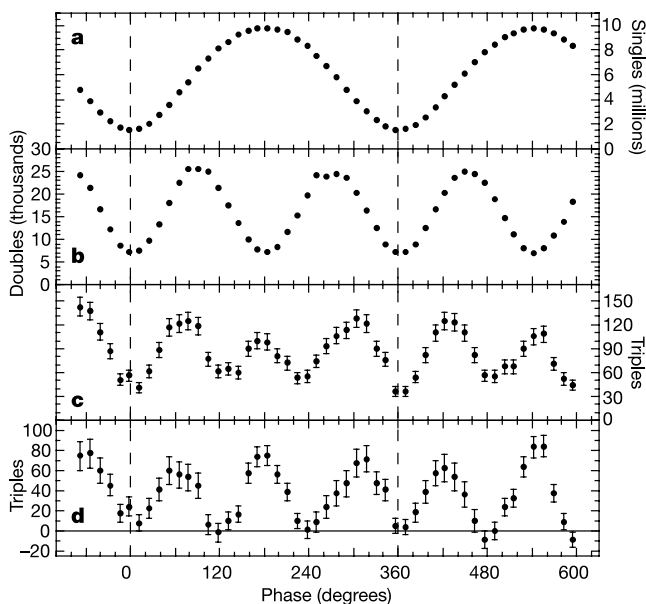


Figure 2 Super-resolving phase measurement with two and three photons. The polarization analyser is configured with one detector on the -45° output and two on $+45^\circ$. The input, a combination of down-converted and laser photons, is constructed to give the state $|3 :: 0\rangle_{H,V}^{3\phi}$ upon post-selection (see text). All graphs show detections per 30-s counting interval as the phase ϕ is changed by translating a phase-shifter prism. **a**, Singles detection at -45° , that is, detection of $|0, 1\rangle_{\pm 45^\circ}$, shows oscillation with ϕ . **b**, Two-fold coincidence detection of $|1, 1\rangle_{\pm 45^\circ}$ shows oscillation with 2ϕ . **c**, Three-fold coincidence detection of $|2, 1\rangle_{\pm 45^\circ}$ shows oscillation with 3ϕ . Error bars indicate $\pm\sigma$ statistical uncertainty. **d**, Three-fold coincidence after background subtraction. Error bars show $\pm\sigma$ statistical uncertainty plus a systematic uncertainty in the background. Dashed vertical bars indicate one full cycle.

Because the DC and LO photons are indistinguishable, there are several different 'paths' to the single outcome $|2, 1\rangle_{\pm 45^\circ}$. For example, the LO photon could go to the -45° detector and the two DC photons to $+45^\circ$. This path interferes with other permutations to give the signal shown in Fig. 2c. A cosine curve fitted to these data shows a visibility of $42 \pm 3\%$. This visibility is unambiguous evidence of indistinguishability and entanglement among the three photons. A non-entangled state of three distinguishable photons could also show threefold coincidence oscillation at 3ϕ , but with a maximal visibility of 20%.

Figure 2d shows the same triples data after subtraction of background from accidental triples. In addition to the signal of interest from 1 DC pair + 1 LO photon, we also see events from 2 DC pairs, from 3 LO photons, and from 2 LO photons + 1 DC pair. We calculate these backgrounds from independent measurements of single and double detection rates for the DC and LO sources alone. Coincidence background is calculated by the statistics of uncorrelated sources using a time-window of 12.5 ns, the laser pulse period. Incoherence of the various contributions is ensured by sweeping the path length of the LO photon over $\pm 2 \mu\text{m}$ during acquisition. The calculated background has some variation with ϕ , so it is important to note that it is qualitatively different from the observed 3ϕ signal. Per 30-s interval, the accidental background contributes 22 ± 1 as a constant component, an average of 23 ± 1 oscillating with 2ϕ , 4 ± 1 with 1ϕ and <1 with 3ϕ . Here and elsewhere, uncertainty in the counting circuitry's dead-time introduces a systematic error.

It is also possible to see 3ϕ behaviour detecting a single polarization (Fig. 3). This measurement corresponds to the original proposals for atomic spectroscopy and lithography^{4,12}. It gives a far weaker signal, in part because the maximum overlap of the state $|3 :: 0\rangle_{H,V}^{3\phi}$ with $|3, 0\rangle_{\pm 45^\circ}$ is smaller than with $|2, 1\rangle_{\pm 45^\circ}$. Also, the chance that all three photons go to distinct detectors (as needed for coincidence detection) is smaller for $|3, 0\rangle_{\pm 45^\circ}$. With these limitations, we are able to see the 3ϕ behaviour, but only by subtracting a considerable coincidence background.

Using linear optical elements and post-selection, we have constructed a multiparticle entangled state useful for super-resolving phase measurements. The demonstrated resolution is not only better than for a single infrared photon, it is better than could be achieved with any single photon in the experiment, including the down-conversion pump photons. Given the difficulty of generating coherent short-wavelength sources, this is encouraging for the

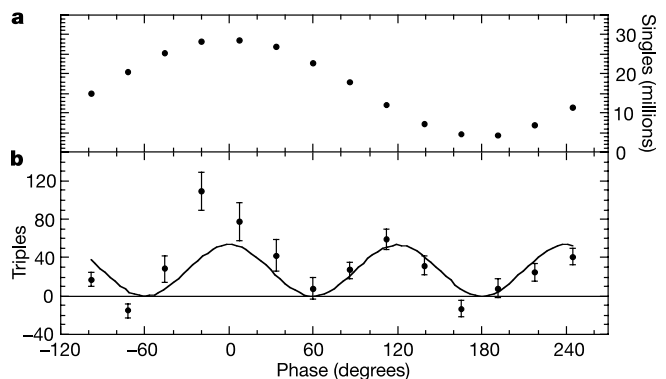


Figure 3 Super-resolving phase measurement with a single detected polarization. The polarization analyser is configured to detect $|3, 0\rangle_{\pm 45^\circ}$, that is, with three detectors on the $+45^\circ$ channel. The input state is the same as in Fig. 2. Graphs show detections per 300-s counting interval as the phase ϕ is changed. **a**, Singles detection shows oscillation with ϕ . **b**, Three-fold coincidence detection (after background subtraction) shows oscillation with 3ϕ . Error bars show $\pm\sigma$ statistical uncertainty plus a systematic uncertainty in the background. Curve is the expected signal $A[1 + \cos(3\phi)]$ with A chosen for best fit.

prospects of proposals such as quantum-interferometric optical lithography. The method can be adapted to generate entangled states of arbitrarily large photon number. Because prior entanglement is not required, the procedure would work well with single-photon-on-demand sources^{29,30}, which promise to be more efficient and scalable than down-conversion sources. Scalability would also be enhanced by the use of photon-number-resolving detectors. The construction proceeds from spatially separated, unentangled photons to a maximally entangled state in a single spatial mode, a state suitable for Heisenberg-limited phase measurements. □

Received 22 December 2003; accepted 16 March 2004; doi:10.1038/nature02493.

- Barish, B. C. & Weiss, R. LIGO and the detection of gravitational waves. *Phys. Today* **52**, 44–50 (1999).
- Caron, B. *et al.* The Virgo interferometer. *Class. Quantum Gravity* **14**, 1461–1469 (1997).
- Holland, M. J. & Burnett, K. Interferometric detection of optical-phase shifts at the Heisenberg limit. *Phys. Rev. Lett.* **71**, 1355–1358 (1993).
- Bollinger, J. J., Itano, W. M., Wineland, D. J. & Heinzen, D. J. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A* **54**, R4649–R4652 (1996).
- Dowling, J. P. Correlated input-port, matter-wave interferometer: Quantum-noise limits to the atom-laser gyroscope. *Phys. Rev. A* **57**, 4736–4746 (1998).
- Ou, Z. Y. Fundamental quantum limit in precision phase measurement. *Phys. Rev. A* **55**, 2598–2609 (1997).
- Campos, R. A., Gerry, C. C. & Benmoussa, A. Optical interferometry at the Heisenberg limit with twin Fock states and parity measurements. *Phys. Rev. A* **68**, 023810 (2003).
- D'Angelo, M., Chekhova, M. V. & Shih, Y. Two-photon diffraction and quantum lithography. *Phys. Rev. Lett.* **87**, 013602 (2001).
- Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Rauschenbeutel, A. *et al.* Step-by-step engineered multiparticle entanglement. *Science* **288**, 2024–2028 (2000).
- Zhao, Z. *et al.* Experimental violation of local realism by four-photon Greenberger-Horne-Zeilinger entanglement. *Phys. Rev. Lett.* **91**, 180401 (2003).
- Kok, P., Lee, H. & Dowling, J. P. Creation of large-photon-number path entanglement conditioned on photodetection. *Phys. Rev. A* **65**, 052104 (2002).
- Fiurasek, J. Conditional generation of n-photon entangled states of light. *Phys. Rev. A* **65**, 053818 (2002).
- Pryde, G. J. & White, A. G. Creation of maximally entangled photon-number states using optical fiber multiports. *Phys. Rev. A* **68**, 052315 (2003).
- Hofmann, H. F. Generation of highly non-classical n-photon polarization states by super-bunching at a photon bottleneck. Preprint at (<http://arxiv.org/quant-ph/0311198>) (2003).
- Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
- Franson, J. D., Donegan, M. M., Fitch, M. J., Jacobs, B. C. & Pittman, T. B. High-fidelity quantum logic operations using linear optical elements. *Phys. Rev. Lett.* **89**, 137901 (2002).
- Resch, K. J., Lundeen, J. S. & Steinberg, A. M. Conditional-phase switch at the single-photon level. *Phys. Rev. Lett.* **89**, 037904 (2002).
- Mitchell, M. W., Ellenor, C. W., Schneider, S. & Steinberg, A. M. Diagnosis, prescription, and prognosis of a Bell-state filter by quantum process tomography. *Phys. Rev. Lett.* **91**, 120402 (2003).
- O'Brien, J. L., Pryde, C. J., White, A. G., Ralph, T. C. & Branning, D. Demonstration of an all-optical quantum controlled-NOT gate. *Nature* **426**, 264–267 (2003).
- Sanders, B. C., Milburn, G. J. & Zhang, Z. Optimal quantum measurements for phase-shift estimation in optical interferometry. *J. Mod. Opt.* **44**, 1309–1320 (1997).
- Huelga, S. F. *et al.* Improvement of frequency standards with quantum entanglement. *Phys. Rev. Lett.* **79**, 3865–3868 (1997).
- Rarity, J. G. *et al.* 2-photon interference in a Mach-Zehnder interferometer. *Phys. Rev. Lett.* **65**, 1348–1351 (1990).
- Ou, Z. Y., Zou, X. Y., Wang, L. J. & Mandel, L. Experiment on nonclassical 4th-order interference. *Phys. Rev. A* **42**, 2957–2965 (1990).
- Meyer, V. *et al.* Experimental demonstration of entanglement-enhanced rotation angle estimation using trapped ions. *Phys. Rev. Lett.* **86**, 5870–5873 (2001).
- Rarity, J. C. & Tapster, P. R. Three-particle entanglement from entangled photon pairs and a weak coherent state. *Phys. Rev. A* **59**, R35–R38 (1999).
- Rarity, J. G., Tapster, P. R. & Loudon, R. Nonclassical interference between independent sources. Preprint at (<http://arxiv.org/quant-ph/9702032>) (1997).
- Takeuchi, S. Beamlike twin-photon generation by use of type II parametric downconversion. *Opt. Lett.* **26**, 843–845 (2001).
- Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
- Solomon, G. S., Pelton, M. & Yamamoto, Y. Single-mode spontaneous emission from a single quantum dot in a three-dimensional microcavity. *Phys. Rev. Lett.* **86**, 3903–3906 (2001).

Acknowledgements We thank K. Resch and J. O'Brien for discussions, and J. Dowling and D. R. Schumliar for inspiration. This work was supported by the National Science and Engineering Research Council of Canada, Photonics Research Ontario, the Canadian Institute for Photonic Innovations and the DARPA-QuIST program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W.M. (mitchell@physics.utoronto.ca).

Increased seasonality in Middle East temperatures during the last interglacial period

Thomas Felis^{1,2}, Gerrit Lohmann^{1,2}, Henning Kuhnert², Stephan J. Lorenz³, Denis Scholz⁴, Jürgen Pätzold^{1,2}, Saber A. Al-Rousan⁵ & Salim M. Al-Moghrabi^{5*}

¹DFG Forschungszentrum Ozeanränder, ²Fachbereich Geowissenschaften, Universität Bremen, 28359 Bremen, Germany

³Max-Planck-Institut für Meteorologie, Modelle & Daten, 20146 Hamburg, Germany

⁴Heidelberger Akademie der Wissenschaften, 69120 Heidelberg, Germany

⁵Marine Science Station, University of Jordan & Yarmouk University, 77110 Aqaba, Jordan

* Present address: Aqaba Special Economic Zone Authority, 77110 Aqaba, Jordan

The last interglacial period (about 125,000 years ago) is thought to have been at least as warm as the present climate¹. Owing to changes in the Earth's orbit around the Sun, it is thought that insolation in the Northern Hemisphere varied more strongly than today on seasonal timescales², which would have led to corresponding changes in the seasonal temperature cycle³. Here we present seasonally resolved proxy records using corals from the northernmost Red Sea, which record climate during the last interglacial period, the late Holocene epoch and the present. We find an increased seasonality in the temperature recorded in the last interglacial coral. Today, climate in the northern Red Sea is sensitive to the North Atlantic Oscillation^{4,5}, a climate oscillation that strongly influences winter temperatures and precipitation in the North Atlantic region. From our coral records and simulations with a coupled atmosphere–ocean circulation model, we conclude that a tendency towards the high-index state of the North Atlantic Oscillation during the last interglacial period, which is consistent with European proxy records^{6–8}, contributed to the larger amplitude of the seasonal cycle in the Middle East.

The Arctic Oscillation/North Atlantic Oscillation (AO/NAO), the Northern Hemisphere's dominant mode of atmospheric variability, exerts a strong influence on mid- and high-latitude continental climate by modulating the strength of the subpolar westerlies at interannual to interdecadal timescales^{9,10}. Previous work has shown that the northernmost Red Sea represents a location to study past AO/NAO-related atmospheric variability over the Northern Hemisphere, and that annually banded corals from this subtropical site provide proxy records of this variability over the past centuries^{4,5}. This narrow, desert-enclosed ocean basin is influenced by mid-latitude continental climate^{5,11} and is sensitive to atmospheric processes owing to a weak water column stratification¹².

Two fossil coral colonies (*Porites*) were collected near Aqaba on the Jordanian coast of the Gulf of Aqaba, the northeastern extension of the northernmost Red Sea (Fig. 1a). Colony AQB-10-B was recovered from a canal cut into the modern reef flat, whereas colony AQB-3-A was collected from a complex of raised reef terraces. X-radiographs revealed annual density bands and were used to identify areas within the colonies that appear to be unaffected by diagenetic alteration (Fig. 1b, c). X-ray diffraction analyses of these areas indicate an aragonite content of 98–99%, and petrographic thin sections show only traces of secondary aragonite. The bimonthly resolution time series of both coral $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ (not shown) generated from these areas show clear annual cycles, suggesting that the sampled sections were not subject to major diagenetic alterations with respect to stable isotopes (see Methods and Supplementary Information).

Radiocarbon dating indicates that coral AQB-10-B grew 2.9 kyr ago during the late Holocene. The age of coral AQB-3-A is 121.9 (+7.0/−6.3) kyr, based on U-series dating including a correction for open-system behaviour. Additional uncertainty due to model assumptions is reflected in the larger age error compared to usual U-series dating of corals¹³ (see Methods). The latter coral grew during the last interglacial period, which, in the Middle East, is documented between 124 and 119 kyr ago based on a U-series dated speleothem record of eastern Mediterranean climate¹⁴, with the main peak at 122 kyr coinciding with the coral's age.

The late Holocene and last interglacial corals provide bimonthly resolution $\delta^{18}\text{O}$ time series for time windows of 98 and 44 yr, respectively. Multitaper method spectral analysis reveals significant variance at interannual periods of 5–6 yr in both records (Fig. 2c, d), which can be interpreted as an indication of AO/NAO-like atmospheric variability over the Northern Hemisphere at 2.9 and 122 kyr ago. Similar variability is evident in the time series of a modern coral from the northernmost Red Sea (Fig. 2a), where it is strongly linked

with regional sea surface temperature (SST) and the AO/NAO⁴. Cross-spectral analysis reveals that this interannual variability is highly coherent and in phase with the AO index (Fig. 2b), although a minor fraction is associated with weaker and non-stationary tropical Pacific teleconnections modulated by higher-latitude atmospheric circulation^{4,11}. Although the AO/NAO is most pronounced during winter, it is present throughout the year^{15,16}. Consequently, spectral analyses were performed with bimonthly resolution time series, reflecting variability throughout the year. This procedure is supported by the finding that the physical mechanism that provides a link between the AO/NAO and Middle East climate during winter⁵ is similar to that for interannual variability throughout the year (see Supplementary Fig. S1): a high-pressure anomaly over the Mediterranean Sea associated with the AO/NAO favours an anticyclonic flow of surface winds in the eastern Mediterranean, which results in advection of colder air from southeastern Europe, controlling SST and coral $\delta^{18}\text{O}$ variability in the northern Red Sea⁵.

The most striking feature of the coral $\delta^{18}\text{O}$ time series is increased seasonality in the last interglacial record compared to the modern and late Holocene records (Fig. 2). Because coral $\delta^{18}\text{O}$ is influenced by both temperature and $\delta^{18}\text{O}$ of sea water, we applied the coral Sr/Ca palaeothermometer to the fossil corals and to three modern reference corals. Combined Sr/Ca and $\delta^{18}\text{O}$ analyses on modern corals show that both proxies satisfactorily document the seasonal

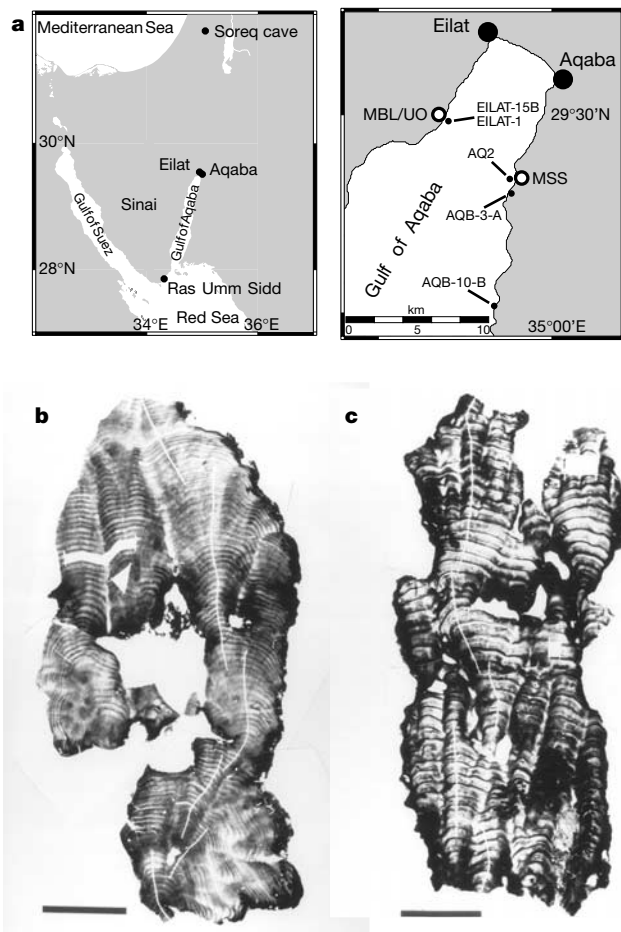


Figure 1 Maps of the northernmost Red Sea, and X-radiographs of the two fossil *Porites* corals. **a**, The sites of coral collection in the northern Gulf of Aqaba. Coral AQB-10-B was collected at 34° 57.84' E, 29° 22.91' N; coral AQB-3-A was collected at 34° 58.28' E, 29° 27.12' N. MBL/UO, H. Steinitz Marine Biology Laboratory/Underwater Observatory, Eilat; MSS, Marine Science Station, Aqaba; sea surface temperature was measured at UO²⁵. **b**, **c**, X-radiograph positive prints of 5-mm-thick slabs sliced parallel to the growth axis of coral AQB-10-B (**b**; 2.9 kyr) and coral AQB-3-A (**c**; 122 kyr). Alternating bands of high (dark colour) and low skeletal density (light colour) are visible. One year is represented by a high-density/low-density band pair. The sampling transect appears as a white line. The corals are about 60 cm high and 25–30 cm in diameter. Scale bars, 10 cm.

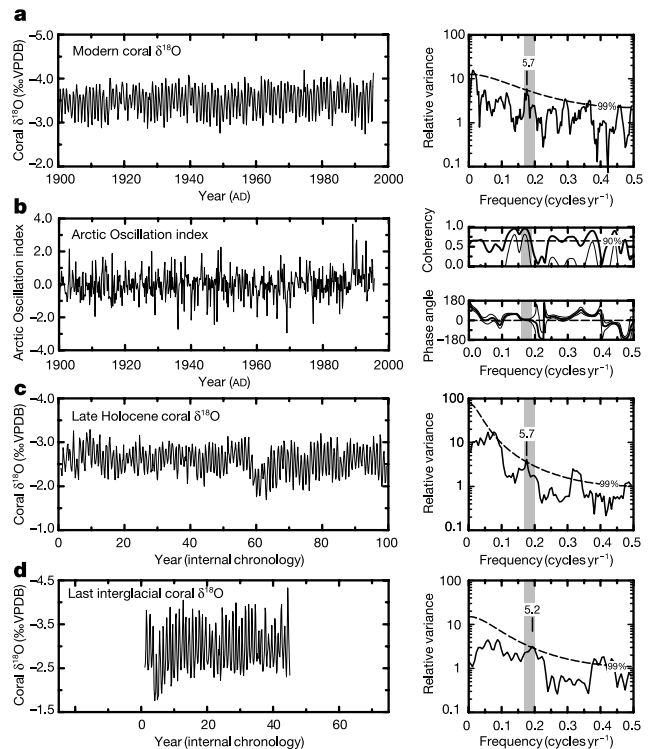


Figure 2 Time series of coral $\delta^{18}\text{O}$ based on modern, late Holocene, and last interglacial *Porites* colonies from the northernmost Red Sea and their spectral properties. Bimonthly coral $\delta^{18}\text{O}$ time series (left panel) and results of multitaper method spectral analysis with red noise null hypothesis³⁰ (number of tapers, 3; bandwidth parameter, 2; 99% significance level is indicated) (right panel) for **a**, a modern coral from Ras Umm Sidd⁴ (RUS-95, AD 1750–1995), **c**, a late Holocene coral (AQB-10-B, 2.9 kyr) and **d**, a last interglacial coral (AQB-3-A, 122 kyr) from Aqaba. **b**, Bimonthly time series of the Arctic Oscillation (AO) index¹⁰ (left panel). Cross-spectral analysis between the time series of the modern coral and the AO index (right panel). The 90% confidence level for coherence is indicated. Spectral analyses were performed for bimonthly, detrended and normalized time series with the average seasonal cycle removed.

SST cycle of 5.4 °C, indicating a seasonal cycle between 4.5 and 5.6 °C (Fig. 3a), in agreement with earlier findings that $\delta^{18}\text{O}$ seasonality in northern Red Sea corals is mainly controlled by temperature⁴. Both Sr/Ca and $\delta^{18}\text{O}$ of the late Holocene coral indicate a seasonal SST cycle of 5.2 °C at 2.9 kyr ago (Fig. 3b), similar to today. In contrast, in the last interglacial coral both proxies indicate increased SST seasonality of 8.4 °C at 122 kyr ago (Fig. 3d) (see Methods and Supplementary Information).

In order to understand the physical mechanisms responsible for increased SST seasonality in the northernmost Red Sea during the last interglacial, and seasonality similar to today at 2.9 kyr ago, we performed coupled atmosphere–ocean general circulation model simulations (ECHO-G) for last interglacial, late Holocene, pre-industrial, and modern conditions (see Methods). Consistent with coral-based results, the corresponding modelled SST indicates increased seasonality during the last interglacial, and seasonality similar to modern and pre-industrial conditions at 3 kyr ago (Fig. 3c, e). Consistent with model-based SST, the modelled Middle East surface air temperature (SAT) anomalies indicate warmer

summers and colder winters relative to modern conditions during the last interglacial (Fig. 4a, b). This increased seasonality would usually be explained by an amplified seasonal insolation cycle at that time (see Supplementary Fig. S4).

Indeed, our model suggests that warmer Middle East summers during the last interglacial result from increased summer insolation, as they are part of a spatially homogenous warming pattern over mid-latitude continental areas where insolation was enhanced (Fig. 4b). However, the model suggests that colder Middle East winters at that time did not solely result from reduced winter insolation at these latitudes, but are associated with the AO/NAO. The modelled winter SAT difference between last interglacial and modern climate reveals a warming and cooling pattern over the North Atlantic and adjacent continental areas that cannot be explained by differences in direct insolation forcing, but that resembles the spatial signature of the AO/NAO^{9,10}. This winter SAT anomaly indicates a tendency towards the AO/NAO high-index state during the last interglacial, with warmer winters in central Europe owing to increased advection of warm oceanic air

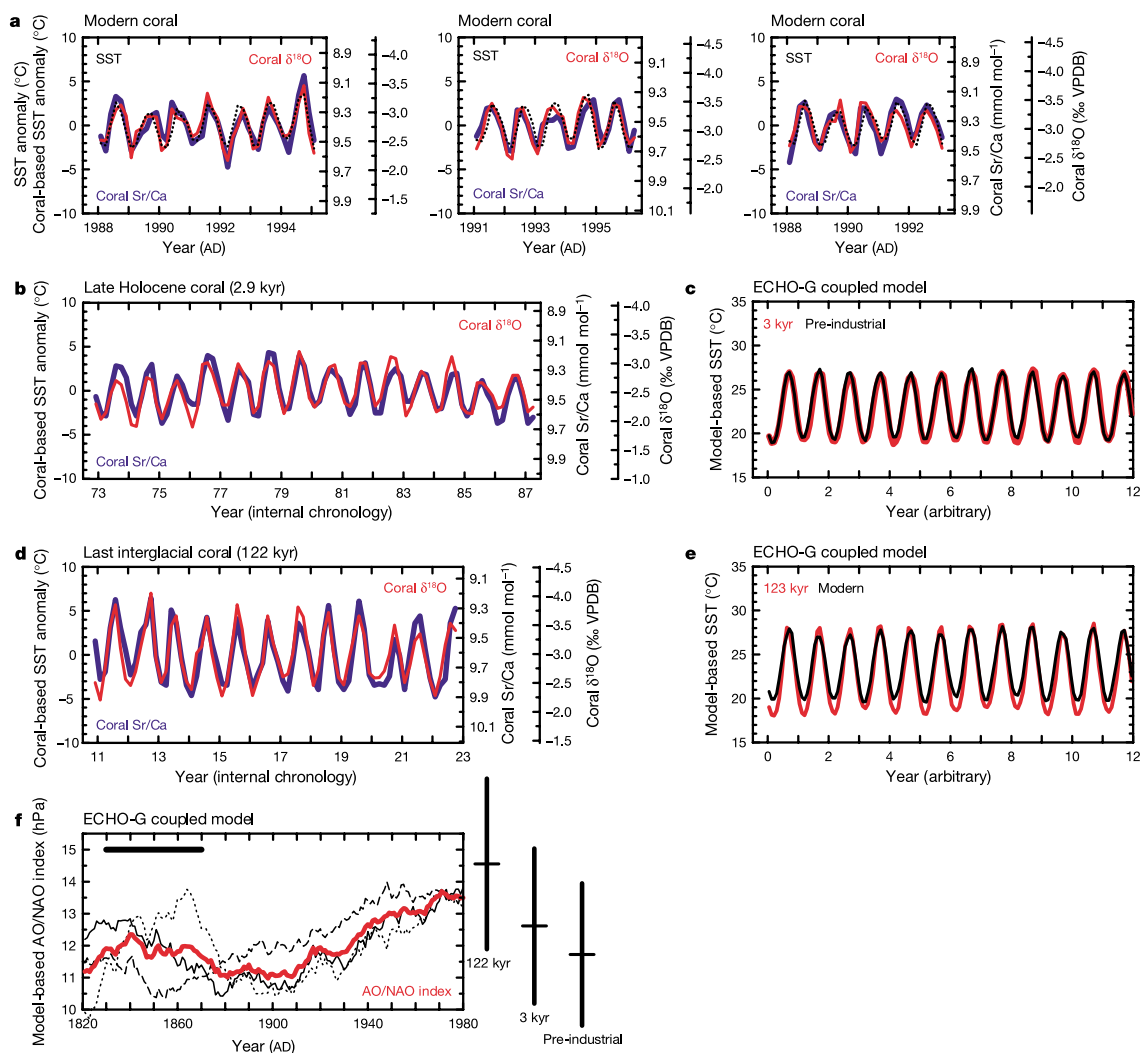


Figure 3 Coral-based sea surface temperature (SST) anomalies for the northernmost Red Sea and ECHO-G model-based SST and AO/NAO indices. Bimonthly *Porites* coral $\delta^{18}\text{O}$ (red) and Sr/Ca (blue) time series and coral-based SST anomalies (respective mean was subtracted). **a**, Modern corals EILAT-15B (left), AQ2 (centre), EILAT-1 (right), *in situ* SST²⁵ (dotted line); **b**, late Holocene (AQB-10-B; 2.9 kyr) and **d**, last interglacial coral (AQB-3-A; 122 kyr). **c**, **e**, Modelled monthly SST index of the coral region for 3 kyr (**c**, red),

123 kyr (**e**, red), pre-industrial (around AD 1830; **c**, black) and modern conditions (around AD 1980; **e**, black). **f**, Modelled AO/NAO index (December–February). Three greenhouse gas scenarios (black lines) and the ensemble mean (red line) for the period AD 1800–2000 (centred 41-yr running means). Mean standard deviations for 41 winters centred at 122 kyr, 3 kyr and the pre-industrial period (thick horizontal bar) are shown on the right. See Supplementary Fig. S7 for index definitions.

from the west, and colder winters in the Middle East owing to increased advection of cold continental air from the north (Fig. 4a). This implies that the AO/NAO contributed to increased SST seasonality in the northernmost Red Sea during the last interglacial through winter cooling. This is consistent with (1) coral-based results of AO/NAO-like interannual variability during the last

interglacial, (2) observations during recent decades where a shift towards the high-index AO/NAO is accompanied by colder winters in the northernmost Red Sea leading to increased seasonality, and (3) a strong relationship between interannual variability of modelled regional winter SST and the AO/NAO (see Supplementary Figs S5, S6). Furthermore, a tendency towards the high-index AO/NAO provides an explanation for warmer winters in central Europe during the last interglacial, as indicated by terrestrial proxy climate records^{6–8}, a finding in conflict with an explanation via reduced winter insolation at these latitudes.

Consistent with coral- and model-based SST, the modelled northernmost Red Sea SAT anomalies indicate seasonality similar to modern and pre-industrial conditions at 3 kyr ago (Fig. 4c, d). The winter SAT anomaly indicates a tendency towards the high-index AO/NAO at 3 kyr ago relative to pre-industrial conditions, which is less pronounced compared to the last interglacial. Compared to the latter, the amplified seasonal insolation cycle at 3 kyr ago is also less pronounced (see Supplementary Fig. S4). For the last interglacial, the winter SAT anomaly from pre-industrial climate indicates a more pronounced tendency towards the high-index AO/NAO compared to that from modern conditions (see Supplementary Fig. S8).

The modelled SAT anomalies, as well as the modelled AO/NAO indices, suggest a combined response of the AO/NAO to seasonal insolation changes on orbital timescales and to atmospheric greenhouse gases (Fig. 3f). Consistent with other model-based studies^{17,18}, a greenhouse gas increase from pre-industrial to modern values results in a slight tendency towards the high-index AO/NAO. The pronounced last interglacial high-index AO/NAO relative to 3 kyr ago and pre-industrial climate, however, can only be explained by differences in insolation forcing, as greenhouse gas concentrations were similar. Interestingly, the interannual AO/NAO variability is nearly unaffected by the orbital forcing, consistent with coral-based results. The physical mechanism linking orbital forcing and the last interglacial high-index AO/NAO most probably involves reduced boreal winter insolation in the tropics (Supplementary Fig. S4). In the model, this leads to a reduced pole-to-equator temperature gradient, and a subsequent weakening of the Hadley cell accompanied by planetary wave activity, with increased Icelandic low and subtropical North Atlantic/eastern Mediterranean highs. The anomalous circulation pattern represents a quasi-equilibrium response to thermal forcing linked to land–sea temperature contrasts and orography¹⁹. Furthermore, a northward shift of the North American and Atlantic jet stream by downward propagating stratospheric anomalies is consistent with the high-index AO/NAO¹⁰.

The AO/NAO, the dominant mode of Northern Hemisphere climate variability on interannual to interdecadal timescales^{9,10}, has also been suggested to be important on millennial timescales^{20,21}. In addition, our approach (of combining seasonal resolution coral proxy records from a climatically sensitive, exceptionally northern, subtropical reef site with coupled atmosphere–ocean circulation

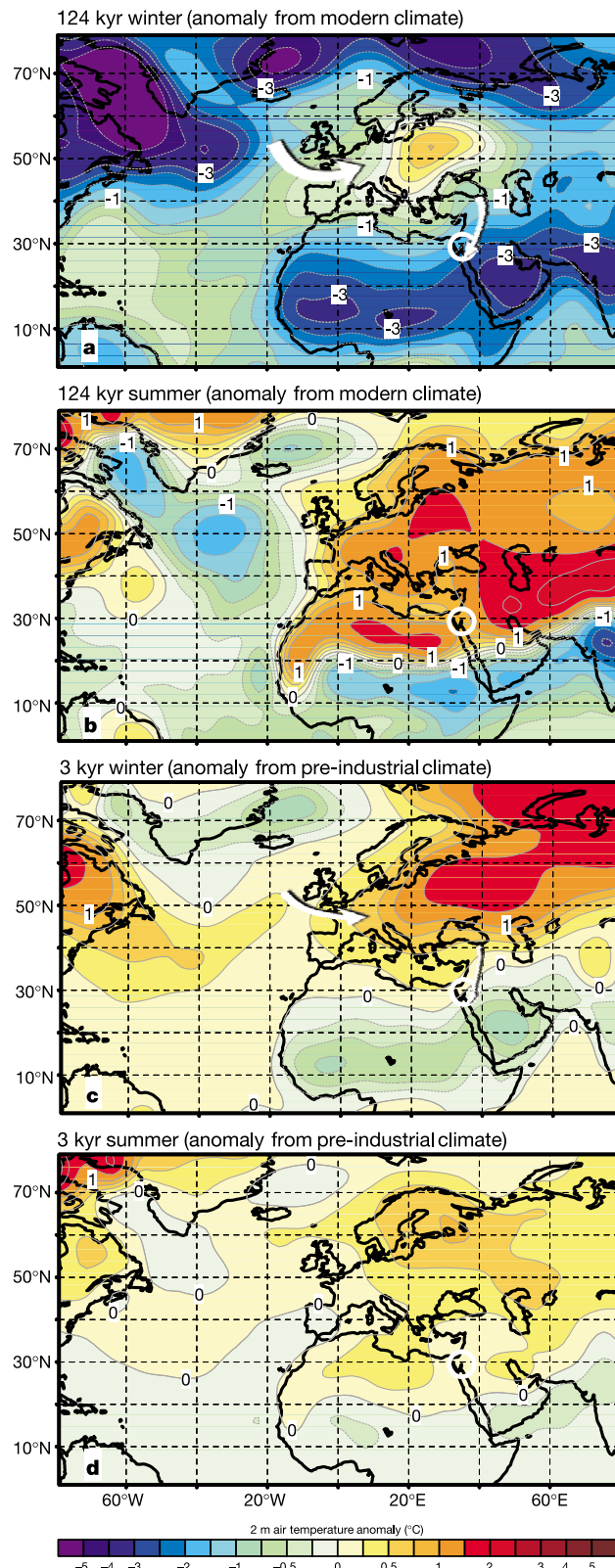


Figure 4 Near-surface air temperature anomalies for the last interglacial and the late Holocene based on the coupled atmosphere–ocean general circulation model ECHO-G. Difference between last interglacial (124 kyr) and modern climate (AD 1975–85) for **a**, winter (December, January, February; DJF) and **b**, summer (June, July, August; JJA). The corresponding anomalies from pre-industrial climate are shown in Supplementary Fig. S8. Difference between late Holocene (3 kyr) and pre-industrial climate (AD 1820–50) for **c**, winter (DJF) and **d**, summer (JJA). The corresponding anomalies from modern climate are shown in Supplementary Fig. S9. Near surface wind anomaly is schematically represented as white arrows. An average of 11 simulation years has been applied to the last interglacial and late Holocene climate centred at the respective time period. The region of coral collection in the northernmost Red Sea is marked by a white circle.

model simulations) suggests an important role of the AO/NAO in modulating regional Northern Hemisphere climate patterns and seasonality on orbital timescales. The cross-validation of well-dated, high-resolution palaeoclimatic records and state-of-the-art climate models provides a strong tool for evaluating the sensitivity of different modes of climate variability to natural and anthropogenic forcing factors. This provides a crucial step in understanding and predicting pronounced changes in past, present and future climate. □

Methods

Microsampling, oxygen isotope and Sr/Ca analyses, age model

Microsampling, $\delta^{18}\text{O}$ analyses, age model construction, and interpolation to a bimonthly resolution were performed as recently described for a modern coral from the northernmost Red Sea⁴. The microdrill bit diameter was adapted to a coral's mean growth rate as estimated from X-radiographs in order to obtain at least six samples per year on average by continuous spot-sampling. For Sr/Ca analyses, inductively coupled plasma mass spectrometry was used (see Supplementary Information).

Radiocarbon and U-series dating

An accelerator mass spectrometry (AMS) ^{14}C date of $3,290 \pm 35$ yr before present (BP) was determined on coral AQB-10-B (KIA13708) at the Leibniz-Labor for Radiometric Dating and Isotope Research (Kiel, Germany). The ^{14}C age was converted to calendar age using the CALIB 4.3 calibration program²², yielding an age of 2,910 calibrated yr BP (2σ range: 3,002–2,807 cal. yr BP). A ΔR value of 154 yr was used to correct for regional differences in reservoir age, based on six AMS ^{14}C dates determined on two modern coral cores with established age models and data from the Marine Reservoir Correction Database (see Supplementary Table S1). Six thermal ionization mass spectrometry (TIMS) U-series dates were determined on coral AQB-3-A at the Forschungsstelle für Radiometrische Altersbestimmungen of the Heidelberger Akademie der Wissenschaften (Heidelberg, Germany).

The U-series ages range from 137.3 to 126.4 kyr ago and the initial $\delta^{234}\text{U}$ values are elevated, ranging from 215‰ to 385‰ and suggesting open-system behaviour of U-series isotopes. In order to solve this problem, a model approach was applied that yields an isochron age of 121.9 ($+7.0/-6.3$) kyr (2σ range)¹³. The larger age error compared to usual U-series dating of corals arises from additional uncertainty due to the model assumptions. The age is consistent with peak sea-level conditions during the last interglacial, based on dated coral reef terraces from the Red Sea²³ (136 to 118 kyr ago) and elsewhere²⁴ (128 to 121 kyr ago), as well as with the last interglacial period in the Middle East¹⁴ (124 to 119 kyr ago). Moreover, coral AQB-3-A was collected from a complex of raised reef terraces with an elevation in the range of other last interglacial reef terraces along the Red Sea coast²³.

Calibration of the proxies

The seasonal maxima and minima in the Sr/Ca record of a modern coral (EILAT-15B) were tied to the corresponding extreme values in a monthly record of *in situ* SST²⁵. A linear least-squares regression was then carried out for bimonthly interpolated Sr/Ca and SST data (with SST defined as the independent variable), giving a relationship of: $\text{Sr/Ca} \times 10^3 = 10.781(\pm 0.1181) - 0.0597(\pm 0.00501) \times \text{SST}$ ($r^2 = 0.78$). The slope of this regression equation is similar to that of calibrations at other locations²⁶. The same procedure was applied to the coral $\delta^{18}\text{O}$ record of EILAT-15B, giving a relationship of: $\delta^{18}\text{O} = 0.801(\pm 0.2773) - 0.1514(\pm 0.01176) \times \text{SST}$ ($r^2 = 0.81$). The slope is similar to that of a calibration from the region⁴.

Using these equations to convert coral Sr/Ca and $\delta^{18}\text{O}$ to SST reveals large offsets in the coral-based mean SST between coral EILAT-15B and two other modern corals (AQ2, EILAT-1) for both proxies, most probably due to so-called vital effects. We therefore do not interpret coral Sr/Ca and $\delta^{18}\text{O}$ in terms of absolute SST, but quantitative estimates of the range of the seasonal SST cycle are possible, as the slope of the proxy-SST calibrations remains unaffected. Modern coral $\delta^{18}\text{O}$ data are from ref. 27, age models were constructed according to ref. 4.

Global circulation model and experimental set-up

The coupled atmosphere-ocean general circulation model ECHO-G is applied²⁸. The atmospheric part of ECHO-G is the general circulation model ECHAM4 with its T30 resolution, which corresponds to a gaussian longitude-latitude grid of approximately $3.8^\circ \times 3.8^\circ$. ECHAM4 is coupled to the HOPE ocean model including a dynamic-thermodynamic sea-ice model. ECHO-G was adapted to account for the influence of variations in the annual distribution of solar radiation resulting from the varying orbital parameters²⁹, which were calculated after ref. 2. The timescale of the astronomical forcing was shortened by an acceleration factor of 100 to enable simulations of a >100 kyr period with ECHO-G²⁹. The insolation trends of the last 140 kyr are represented in 1,400 simulation years. Three ensemble experiments for the period 140 kyr ago to AD 1800 were performed with orbital forcing² only. Throughout the experiments, the greenhouse gas concentrations were fixed (latest Holocene values: 280 p.p.m. CO_2 , 700 p.p.b. CH_4 , 265 p.p.b. N_2O) and modern values for vegetation, sea level, and distribution of land, ocean and continental ice were used. The experiments were continued from AD 1800 onward with increasing greenhouse gas concentrations, reaching 370 p.p.m. CO_2 in AD 2000. The temperature anomalies induced by anthropogenic greenhouse gases are shown in Supplementary Fig. S10.

Received 30 January; accepted 5 April 2004; doi:10.1038/nature02546.

- Kukla, G. J. *et al.* Last interglacial climates. *Quat. Res.* **58**, 2–13 (2002).
- Berger, A. L. Long-term variations of daily insolation and Quaternary climatic changes. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
- Montoya, M., von Storch, H. & Crowley, T. J. Climate simulation for 125 kyr BP with a coupled ocean-atmosphere general circulation model. *J. Clim.* **13**, 1057–1072 (2000).
- Felis, T. *et al.* A coral oxygen isotope record from the northern Red Sea documenting NAO, ENSO, and North Pacific teleconnections on Middle East climate variability since the year 1750. *Paleoceanography* **15**, 679–694 (2000).
- Rimbu, N., Lohmann, G., Felis, T. & Pätzold, J. Arctic Oscillation signature in a Red Sea coral. *Geophys. Res. Lett.* **28**, 2959–2962 (2001).
- Zagwijn, W. H. An analysis of Eemian climate in western and central Europe. *Quat. Sci. Rev.* **15**, 451–469 (1996).
- Aalbersberg, G. & Litt, T. Multiproxy climate reconstructions for the Eemian and early Weichselian. *J. Quat. Sci.* **13**, 367–390 (1998).
- Klotz, S., Guiot, J. & Mosbrugger, V. Continental European Eemian and early Würmian climate evolution: comparing signals using different quantitative reconstruction approaches based on pollen. *Glob. Planet. Change* **36**, 277–294 (2003).
- Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
- Thompson, D. W. J. & Wallace, J. M. Regional climate impacts of the Northern Hemisphere Annular Mode. *Science* **293**, 85–89 (2001).
- Rimbu, N., Lohmann, G., Felis, T. & Pätzold, J. Shift in ENSO teleconnections recorded by a northern Red Sea coral. *J. Clim.* **16**, 1414–1422 (2003).
- Eshel, G., Schrag, D. P. & Farrell, B. F. Troposphere-planetary boundary layer interaction and the evolution of ocean surface density: Lessons from Red Sea corals. *J. Clim.* **13**, 339–351 (2000).
- Scholz, D., Mangini, A. & Felis, T. U-series dating of diagenetically altered fossil reef corals. *Earth Planet. Sci. Lett.* **218**, 163–178 (2004).
- Bar-Matthews, M., Ayalon, A. & Kaufman, A. Timing and hydrological conditions of sapropel events in the Eastern Mediterranean, as evident from speleothems, Soreq cave, Israel. *Chem. Geol.* **169**, 145–156 (2000).
- Barnston, A. G. & Livezey, R. E. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Weath. Rev.* **115**, 1083–1126 (1987).
- Rogers, J. C. & McHugh, M. J. On the separability of the North Atlantic oscillation and Arctic oscillation. *Clim. Dyn.* **19**, 599–608 (2002).
- Fyfe, J. C., Boer, G. J. & Flato, G. M. The Arctic and Antarctic Oscillations and their projected changes under global warming. *Geophys. Res. Lett.* **26**, 1601–1604 (1999).
- Shindell, D. T., Miller, R. L., Schmidt, G. A. & Pandolfo, L. Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature* **399**, 452–455 (1999).
- Held, I. M., Ting, M. & Wang, H. Northern winter stationary waves: theory and modeling. *J. Clim.* **15**, 2125–2144 (2002).
- Keigwin, L. D. & Pickart, R. S. Slope water current over the Laurentian Fan on interannual to millennial time scales. *Science* **286**, 520–523 (1999).
- Noren, A. J., Bierman, P. R., Steig, E. J., Lini, A. & Southon, J. Millennial-scale storminess variability in the northeastern United States during the Holocene epoch. *Nature* **419**, 821–824 (2002).
- Stuiver, M. & Reimer, P. J. Extended ^{14}C database and revised CALIB radiocarbon calibration program. *Radiocarbon* **35**, 215–230 (1993).
- Walter, R. C. *et al.* Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* **405**, 65–69 (2000).
- Stirling, C. H., Esat, T. M., Lambeck, K. & McCulloch, M. T. Timing and duration of the Last Interglacial: evidence for a restricted interval of widespread coral reef growth. *Earth Planet. Sci. Lett.* **160**, 745–762 (1998).
- Genin, A., Lazar, B. & Brenner, S. Vertical mixing and coral death in the Red Sea following the eruption of Mount Pinatubo. *Nature* **377**, 507–510 (1995).
- Marshall, J. F. & McCulloch, M. T. An assessment of the Sr/Ca ratio in shallow water hermatypic corals as a proxy for sea surface temperature. *Geochim. Cosmochim. Acta* **66**, 3263–3280 (2002).
- Felis, T., Pätzold, J. & Loya, Y. Mean oxygen-isotope signatures in *Porites* spp. corals: inter-colony variability and correction for extension-rate effects. *Coral Reefs* **22**, 328–336 (2003).
- Legutke, S. & Voss, R. *The Hamburg Atmosphere-Ocean Coupled Circulation Model Echo-G* (Technical Report No. 21, Deutsches Klimarechenzentrum, Hamburg, Germany, 1999).
- Lorenz, S. J. & Lohmann, G. Acceleration technique for Milankovitch type forcing in a coupled atmosphere-ocean circulation model: method and application for the Holocene. *Clim. Dyn.* (submitted).
- Ghil, M. *et al.* Advanced spectral methods for climatic time series. *Rev. Geophys.* **40**, doi:10.1029/2001RG000092 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Segl and her team for stable isotope analyses, A. Abu-Hilal and the Aqaba Marine Science Station for support within the Red Sea Programme, N. Rimbu for discussions, M. Zuther for X-ray diffraction analyses, N. Zatloukal for thin sections, J. Zinke for material, W. Hale for comments, and S. Legutke for support with the ECHO-G model. This work was supported by Bundesministerium für Bildung und Forschung through KIHZ and DEKLIM, and by Deutsche Forschungsgemeinschaft through DFG Research Centre 'Ocean Margins' at Bremen University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.F. (tfelis@allgeo.uni-bremen.de). The data that we report here can be accessed at the World Data Centers for Marine Environmental Sciences (www.wdc-mare.org/) PangaVista?query=@Ref25611) or Paleoclimatology (www.ngdc.noaa.gov/paleo/pubs/felis2004/felis2004.html).

Power requirement of the geodynamo from ohmic losses in numerical and laboratory dynamos

Ulrich R. Christensen¹ & Andreas Tilgner²

¹Max-Planck-Institut für Aeronomie, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany

²Institut für Geophysik, Universität Göttingen, Herzberger Landstrasse 180, 37075 Göttingen, Germany

In the Earth's fluid outer core, a dynamo process converts thermal and gravitational energy into magnetic energy. The power needed to sustain the geomagnetic field is set by the ohmic losses (dissipation due to electrical resistance)¹. Recent estimates of ohmic losses cover a wide range, from 0.1 to 3.5 TW, or roughly 0.3–10% of the Earth's surface heat flow^{1–4}. The energy requirement of the dynamo puts constraints on the thermal budget and evolution of the core through Earth's history^{1–5}. Here we use a set of numerical dynamo models to derive scaling relations between the core's characteristic dissipation time and the core's magnetic and hydrodynamic Reynolds numbers—dimensionless numbers that measure the ratio of advective transport to magnetic and viscous diffusion, respectively. The ohmic dissipation of the Karlsruhe dynamo experiment⁶ supports a simple dependence on the magnetic Reynolds number alone, indicating that flow turbulence in the experiment and in the Earth's core has little influence on its characteristic dissipation time. We use these results to predict moderate ohmic dissipation in the range of 0.2–0.5 TW, which removes the need for strong radioactive heating in the core⁷ and allows the age of the solid inner core to exceed 2.5 billion years.

The limited thermodynamic efficiency of thermal convection means that the power available to drive the dynamo is only a fraction of the total heat flow from the core. Compositional convection, driven by the rejection of light alloying elements from the growing solid inner core, is not limited by Carnot efficiency, but is intimately associated with core cooling. The heat flow from the core must be 5–10 times larger than the ohmic dissipation¹. For high dissipation, the core has to supply a substantial part of the Earth's heat. If this heat flow is due to secular cooling alone, as is conventionally assumed, it implies unrealistic core temperatures early in Earth's history^{1,3}. Significant radiogenic heat production in the core (for example, from decay of ⁴⁰K) would be needed to avoid this. If the core cools rapidly, the solid inner core would have formed only 1 Gyr ago³. It is clearly important to better constrain the actual power requirements of the dynamo.

The observed geomagnetic field could be maintained, in principle, by currents that produce <1 GW of ohmic dissipation², but the actual losses are believed to be much larger. Ohmic dissipation is given by

$$D_{\text{ohm}} = \int (\eta/\mu_0)(\nabla \times \mathbf{B})^2 dV \propto 2\eta E_{\text{mag}}/l_B^2 \quad (1)$$

where η is magnetic diffusivity, μ_0 magnetic permeability, $\mathbf{B}$ magnetic field, E_{mag} magnetic energy and l_B the characteristic length scale of the field. Estimating the ohmic dissipation of the geodynamo suffers from several sources of uncertainty—for example, the magnetic field strength inside the core. More importantly, the scale of the core field is unknown, because magnetization of the Earth's crust shields wavelengths below 3,000 km from observation⁸. Recent geodynamo models can reproduce basic properties of the geomagnetic field, and have been used to estimate ohmic dissipation^{1,2,9,10}. However, the values of some model

parameters are far from Earth-like. In particular, the Ekman number $E = \nu/(\Omega R^2)$, measuring viscous forces relative to rotational forces, and the magnetic Prandtl number $\text{Pm} = \nu/\eta$, are far too large (ν is viscosity, Ω rotation frequency and R core radius). Several models use hyperdiffusivities that suppress small scales in the magnetic and flow fields. A rather low ohmic dissipation of about 0.1–0.3 TW has been estimated when small-scale contributions are ignored^{1,2}. An extrapolation to account for these scales, using the magnetic power spectrum at the core–mantle boundary from a high-resolution dynamo model¹⁰, predicted 1–2 TW of total ohmic dissipation².

Rather than using a single dynamo, we determine the time-average ohmic dissipation in a suite of 24 models, varying the key non-dimensional numbers by factors of 20–30. This allows us to study systematically the dependence on the control parameters and to derive scaling laws that are applicable to the geodynamo. Because the magnetic energy differs substantially between models, we do not

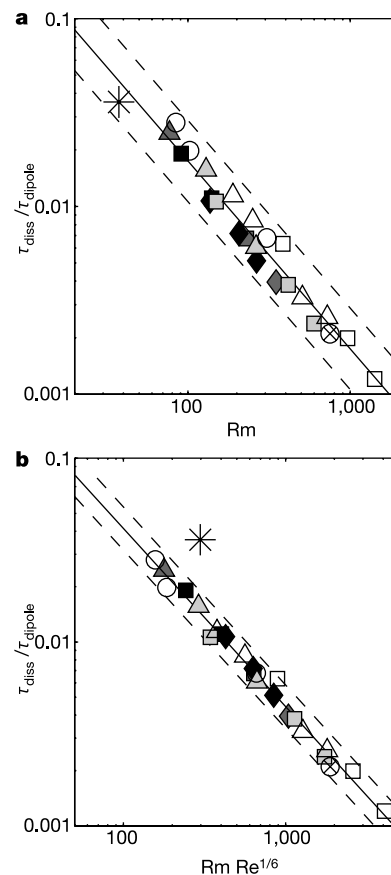


Figure 1 Scaling of magnetic dissipation time. **a**, **b**, Dissipation time in units of dipole decay time versus magnetic Reynolds number (**a**) and versus a combination of magnetic and hydrodynamic Reynolds numbers (**b**). The magnetic Prandtl number is 2–3 for open symbols, 1 for light grey, 0.5 for dark grey and 0.15–0.25 for black symbols. The Ekman numbers are 1.27×10^{-4} (circles), 4.2×10^{-5} (triangles), 1.27×10^{-5} (squares) and 4.2×10^{-6} (diamonds). The reversing dynamo is marked by a circled cross. Best-fitting lines with a slope of -1.0 in **a** and -0.97 in **b** and 3σ limits (broken lines) are drawn. The mean misfit is 16% in **a** and 9% in **b**. All dynamos have been run for at least 40 advection times (R/U), and averages have been taken after the transient adjustment. The numerical resolution varies between 53 and 168 in spherical harmonic degree and 33 and 81 radial points, depending on parameters. At the lowest Ekman number, twofold symmetry in longitude has been assumed; all other cases are for a full sphere. The large asterisk is for the Karlsruhe laboratory dynamo. For the laboratory dynamo the magnetic Reynolds number is based on the cylinder radius R , whereas the hydrodynamic Reynolds number is calculated with the width of the flow ducts $d = 0.05$ m.

scale D_{ohm} directly, but rather scale the magnetic dissipation time:

$$\tau_{\text{diss}} = E_{\text{mag}}/D_{\text{ohm}} \propto I_B^2/2\eta \quad (2)$$

For our models we solve the full magnetohydrodynamic equations without hyperdiffusivities for an incompressible fluid in a rotating spherical shell¹¹. The magnetic fields are dipole-dominated, mostly with stable polarity. We include one case with dipole reversals similar to those in the geomagnetic field¹².

A large-scale magnetic field is converted by nonlinear interaction with the flow field to small scales where it is dissipated. The important parameter for this process is the magnetic Reynolds number $Rm = UR/\eta$, with the r.m.s. velocity U . In Fig. 1a we plot magnetic dissipation time versus Rm and find a simple fit of the form:

$$\tau_{\text{diss}}/\tau_{\text{dipole}} = 1.74 Rm^{-1} \quad (3)$$

Here we normalize with the dipole decay time, $\tau_{\text{dipole}} = R^2/(\pi^2\eta)$, for a full sphere, which is the longest possible time constant for decay of a magnetic field in a stagnant conductor. Cases with lower magnetic Prandtl number (darker shading in Fig. 1a) tend to plot below the fitting line, and those with higher Pm tend to fall above the line. This suggests an additional dependence on Pm , or expressed differently, on the hydrodynamic Reynolds number $Re = Rm/Pm$. A best fit of the form:

$$\tau_{\text{diss}}/\tau_{\text{dipole}} = a(Rm Re^b)^c \quad (4)$$

with $a = 3.58$, $b = 1/6$ and $c = -0.97$, reduces the scatter somewhat (Fig. 1b). We find no clear influence of the Ekman number. The dependence on Re is weak, but to apply equation (4) to the core requires extrapolation over six orders of magnitude in Re , and leads to a factor of ten difference in τ_{diss} compared with equation (3). Adding more free parameters always reduces the misfit, hence one may question if a dependence on Re is really warranted, and if so, whether it also holds for much larger values of the Reynolds number.

In order to resolve this question, we analyse the ohmic dissipation

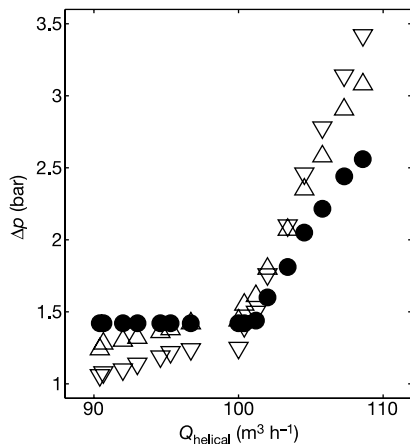


Figure 2 Pressure drop versus flow rate in the Karlsruhe dynamo experiment. Three independent pumps send sodium through disjoint flow loops, two of which are designated as 'helical' and one as 'axial' because of the shape of the path followed by the flow. The cylinder formed by the pipes has a radius $R = 0.95$ m and similar height. Triangles show the pressure drop Δp in the helical loops, and circles that in the axial loop versus flow rate Q_{helical} in the helical loops. The flow in the axial loop is held constant at $112.5 \text{ m}^3 \text{ h}^{-1}$. Above the onset of dynamo action at $Q_{\text{helical}} \approx 100 \text{ m}^3 \text{ h}^{-1}$, the ohmic dissipation is calculated as $D_{\text{ohm}} = \Sigma Q_i(\Delta p_i - \Delta p'_i)$, where the summation is over the three loops. The contribution of viscous friction to the pressure drop Δp^v is obtained by extrapolating Δp from below the threshold of dynamo action. We extrapolate (interpolate) the data to a reference state with $Q = 111 \text{ m}^3 \text{ h}^{-1}$ in all three loops, for which the magnetic field was measured inside the cylinder without recording the pressure drop.

of the Karlsruhe laboratory dynamo^{6,13}, where liquid sodium is pumped through a system of pipes arranged into cells forming a nearly homogeneously conducting cylinder. Here the hydrodynamic Reynolds number is 2.5×10^5 , based on the size of the largest possible turbulent eddies, and the magnetic Prandtl number is 9×10^{-6} , whereas in the models $Re < 500$ and $Pm \geq 0.15$. When the flow rate exceeds a threshold, dynamo action sets in and a sharp rise in the driving pressure drop can be used to calculate the ohmic dissipation (Fig. 2). In order to calculate τ_{diss} , the magnetic energy must be known. We obtain this energy by fitting the magnetic field of a dedicated kinematic dynamo simulation to field measurements performed along the cylinder axis and outside the sodium. A simpler version of this model predicted successfully the onset of dynamo action¹⁴. Finally, we normalize τ_{diss} with the numerically calculated dipole decay time $\tau_{\text{dipole}} = 0.79$ s. The result, marked by an asterisk in Fig. 1, agrees well with the simple scaling on the magnetic Reynolds number alone. The additional dependence on the hydrodynamic Reynolds number under-predicts the dissipation time by a factor of 2.5 (Fig. 1b), which is far outside the estimated uncertainty for the experimental value of 40% and the 3σ limit of the fit to the numerical results. A dependence of τ_{diss} on Re might be plausible, because the small eddies that occur in the flow at high Re

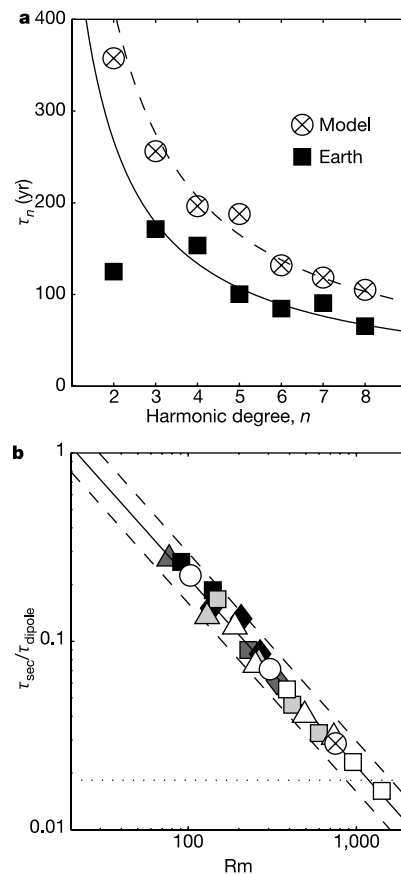


Figure 3 Secular variation time scaling. **a**, Timescale τ_n of secular variation as function of spherical harmonic degree n for the geomagnetic field in the time interval 1840–1990 and, as an example, for the long-term average of the reversing dynamo model. The model data are scaled to real time with $\tau_{\text{dipole}} = 29,000$ yr, obtained for an electrical conductivity $\sigma = \mu_0/\eta = 6 \times 10^5 \text{ S m}^{-1}$ in the core²⁰. Fitting lines of the form $\tau_n = \tau_{\text{sec}}/n$ are included. **b**, Secular variation timescales τ_{sec} versus magnetic Reynolds number. The fitting line is $\tau_{\text{sec}}/\tau_{\text{dipole}} = 21.7 Rm^{-1}$. The dotted horizontal line indicates the Earth value estimated from the fit in **a**, 535 yr in physical units. The predicted $Rm \approx 1,200$ agrees well with the value obtain from estimates of $U = 12\text{--}15 \text{ km yr}^{-1}$ obtained by inverting geomagnetic secular variations for the fluid flow near the core surface²¹.

(low Pm), together with the large-scale flow, may be more efficient in transporting magnetic energy to small scales. However, such an effect will vanish in any case when the turbulent eddies become smaller than the length scale l_B at which diffusion dissipates magnetic energy. One interpretation of our finding is that a weak dependence on the magnetic Prandtl number at values $Pm \approx 1$ disappears for $Pm < 1$. We therefore suggest that the simpler scaling law (equation (3)) represents Earth's core conditions reasonably well.

In order to calculate the dissipation of the geodynamo, we must know R_m and the total magnetic energy in the core. We use the dependence of the secular variation on R_m in our dynamo models to estimate the core value. The timescale of secular variation depends on the spherical harmonic degree n , and is defined as:

$$\tau_n = \left[\frac{\left\langle \sum_{m=0}^n (g_{nm}^2 + h_{nm}^2) \right\rangle}{\left\langle \sum_{m=0}^n (g_{nm}^2 + h_{nm}^2) \right\rangle} \right]^{1/2}$$

where g, h are the Gauss coefficients, the dot marks their time derivative and $\langle \rangle$ the time average. For the geomagnetic field, τ_n decreases with n (ref. 15). To derive a single time constant of secular variation τ_{sec} we attempt a simple fit of the form $\tau_n = \tau_{sec}/n$, although a somewhat stronger dependence on n might better represent the present rate of secular variation (R. Holme, personal communication). Excluding the dipole part, the fit is fair for $n = 2-8$ in the time period 1840–1990 and gives $\tau_{sec} = 535$ yr (Fig. 3a). The secular variation in the dynamo models, averaged over much longer time, follows more closely a $1/n$ -dependence. τ_{sec} depends on the inverse of the magnetic Reynolds number (Fig. 3b). The estimated secular variation time of the geomagnetic field requires $R_m = 1,200$ in the core, which leads to a magnetic dissipation time of 42 yr.

The factor between the mean magnetic field strength inside the model shell and that in degrees n up to 12 on the outer boundary is in the range of 2.5–5 in our non-reversing dynamos and 7.5 in the reversing case. With a likely factor of 5–7.5 for the geodynamo and an r.m.s. field strength ($n < 13$) at the core–mantle boundary of 0.39 mT (ref. 16), we infer 2–3 mT for the field in the core, which gives $E_{mag} = (2.8-6.2) \times 10^{20}$ J. From equation (2) the ohmic dissipation is found to be 0.2–0.5 TW.

For the recently preferred high-power-consumption values of the geodynamo of >1 TW (refs 2, 3, 17), the required heating could be supplied by >200 p.p.m. potassium in the core¹⁷. Although recent experiments suggest that such concentrations are possible^{7,18}, our result suggesting a more moderate power requirement relaxes severe constraints on core evolution, and removes the strong need for heat sources in the core. The inner core could be much older than 1 Gyr; thermal modelling¹ predicts an inner core age of 2.4 Gyr for $D_{ohm} = 0.5$ TW and ~ 3.5 Gyr for $D_{ohm} = 0.2$ TW. As the geodynamo must operate differently in the absence of an inner core or may not operate at all, the existence of a magnetic field of roughly present-day strength over the past 3.5 Gyr (ref. 19) is more easily reconciled with an old inner core. □

Received 3 February; accepted 22 March 2004; doi:10.1038/nature02508.

1. Buffett, B. A. Estimates of heat flow in the deep mantle based on the power requirements of the geodynamo. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014649 (2002).
2. Roberts, P. H., Jones, C. A. & Calderwood, A. R. in *Earth's Core and Lower Mantle* (eds Jones, C. A., Soward, A. M. & Zhang, K.) 100–129 (Taylor & Francis, London, 2003).
3. Labrosse, S. Thermal and magnetic evolution of the Earth's core. *Phys. Earth Planet. Inter.* **140**, 127–143 (2003).
4. Gubbins, D., Alfé, D., Masters, G., Price, D. & Gillan, M. J. Can the Earth's dynamo run on heat alone? *Geophys. J. Int.* **155**, 609–622 (2003).
5. Nimmo, F., Price, G. D., Brodholt, J. & Gubbins, D. The influence of potassium on core and geodynamo evolution. *Geophys. J. Int.* **156**, 363–376 (2004).
6. Stieglitz, R. & Müller, U. Experimental demonstration of the homogeneous two-scale dynamo. *Phys. Fluids* **13**, 561–564 (2001).
7. Gessmann, C. K. & Wood, B. J. Potassium in the Earth's core? *Earth Planet. Sci. Lett.* **200**, 63–78 (2002).
8. Langel, R. A. & Estes, R. H. A geomagnetic field spectrum. *Geophys. Res. Lett.* **9**, 250–253 (1982).
9. Kuang, W. & Bloxham, J. An Earth-like numerical dynamo model. *Nature* **389**, 371–374 (1997).
10. Roberts, P. H. & Glatzmaier, G. A. A test of the frozen-flux approximation using a new geodynamo model. *Phil. Trans. R. Soc. Lond. A* **358**, 1109–1121 (2000).

11. Christensen, U., Olson, P. & Glatzmaier, G. A. Numerical modeling of the geodynamo: a systematic parameter study. *Geophys. J. Int.* **138**, 393–409 (1999).
12. Kutzner, C. & Christensen, U. R. From stable dipolar to reversing numerical dynamos. *Phys. Earth Planet. Inter.* **131**, 29–45 (2002).
13. Müller, U. & Stieglitz, R. The Karlsruhe dynamo experiment. *Nonlin. Proc. Geophys.* **9**, 165–170 (2002).
14. Tilgner, A. Numerical simulation of the onset of dynamo action in an experimental two-scale dynamo. *Phys. Fluids* **14**, 4092–4094 (2002).
15. Hulot, G. & LeMouél, J. L. A statistical approach to the Earth's main magnetic field. *Phys. Earth Planet. Inter.* **82**, 167–183 (1994).
16. Bloxham, J. & Jackson, A. Time-dependent mapping of the magnetic field at the core–mantle boundary. *J. Geophys. Res.* **97**, 19537–19563 (1992).
17. Buffett, B. A. The thermal state of the Earth's core. *Science* **299**, 1675–1676 (2003).
18. Rama Murthy, V., van Westrenen, W. & Fei, Y. Experimental evidence that potassium is a substantial radioactive heat source in planetary cores. *Nature* **423**, 163–165 (2003).
19. McElhinny, M. W. & Senanayake, W. E. Paleomagnetic evidence for the existence of the geomagnetic field 3.5 Ga ago. *J. Geophys. Res.* **85**, 3523–3528 (1980).
20. Secco, R. A. & Schloessin, H. H. The electrical resistivity of solid and liquid Fe at pressures up to 7 GPa. *J. Geophys. Res.* **94**, 5887–5894 (1989).
21. Bloxham, J., Gubbins, D. & Jackson, A. Geomagnetic secular variations. *Phil. Trans. R. Soc. Lond. A* **329**, 415–502 (1989).

Acknowledgements We thank U. Müller for the permission to use unpublished results from the laboratory dynamo experiment. This work was supported by the priority programme “Geomagnetic secular variations” of the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.R.C. (Christensen@linmpi.mpg.de)

Optimal nitrogen-to-phosphorus stoichiometry of phytoplankton

Christopher A. Klausmeier^{1,2}, Elena Litchman^{2,3}, Tanguy Daufresne¹ & Simon A. Levin¹

¹Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA

²School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, Georgia 30332-0230, USA

³Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901, USA

Redfield noted the similarity between the average nitrogen-to-phosphorus ratio in plankton (N:P = 16 by atoms) and in deep oceanic waters (N:P = 15; refs 1, 2). He argued that this was neither a coincidence, nor the result of the plankton adapting to the oceanic stoichiometry, but rather that phytoplankton adjust the N:P stoichiometry of the ocean to meet their requirements through nitrogen fixation, an idea supported by recent modelling studies^{3,4}. But what determines the N:P requirements of phytoplankton? Here we use a stoichiometrically explicit model of phytoplankton physiology and resource competition to derive from first principles the optimal phytoplankton stoichiometry under diverse ecological scenarios. Competitive equilibrium favours greater allocation to P-poor resource-acquisition machinery and therefore a higher N:P ratio; exponential growth favours greater allocation to P-rich assembly machinery and therefore a lower N:P ratio. P-limited environments favour slightly less allocation to assembly than N-limited or light-limited environments. The model predicts that optimal N:P ratios will vary from 8.2 to 45.0, depending on the ecological conditions. Our results show that the canonical Redfield N:P ratio of 16 is not a universal biochemical optimum, but instead represents an average of species-specific N:P ratios.

Laboratory studies show that phytoplankton are flexible in their overall stoichiometry, often matching their nutrient supply at low growth rates^{5–7}. However, this variability is largely due to stored nutrients: underneath these variable pools is the relatively constant composition of the cells' functional machinery⁶. This structural stoichiometry determines the nutrient requirements of the species. Physiological studies have shown that the structural N:P ratio is species-specific⁸. Figure 1 shows the structural N:P ratios of 29 freshwater and marine species (see Supplementary Information). Although the median structural N:P of 17.7 is close to the Redfield ratio, structural N:P ratios range from 7.1 to 43.3, with an outlier at 133.3. Although this data set does not include the structural N:P ratios of the abundant marine picoplankton, *Synechococcus* and *Prochlorococcus*, recent studies^{9,10} have shown that their N:P ratios during exponential growth range from 13.3 to 33.2 and 15.9 to 24.4 respectively, within the range of our data set. We aim to understand the causes of and limits to this interspecific variation, building on recent work on the chemical composition of cellular machinery^{11,12}.

Our approach to deriving optimal phytoplankton stoichiometry has three steps. First, we analyse an underlying model of a single species of phytoplankton given its structural stoichiometry and ecological parameters. Next, we characterize a species by its allocation to two broad classes of cellular machinery: assembly machinery and resource-acquisition machinery. Assembly machinery corresponds to ribosomes, which contain both N and P, and resource-acquisition machinery corresponds to nutrient-uptake proteins and chloroplasts, which contain N but little or no P (refs 6, 12). The allocation strategy determines both the ecological parameters and the structural stoichiometry. We follow the growth-rate hypothesis, that ribosome/RNA content largely explains both the growth rate and the P content of organisms⁶. Finally, we determine the optimal strategy under opposite ecological scenarios: during exponential growth and at competitive equilibrium under different limiting resources (N, P and light). Others have modelled carbon allocation to different cellular components^{13,14}, but these physiological models do not investigate N:P ratios or use competition theory to determine the optimal allocation patterns under diverse ecological conditions.

During exponential growth at saturating resource levels, the optimal strategy maximizes the growth rate, $\mu_{\max}$ (Fig. 2a). Using the parameters given below (see Methods), we find the optimal allocation-to-assembly under exponential growth, $R_a = 0.645$, giving a structural N:P ratio of 8.2. At competitive equilibrium, the optimal strategy minimizes the break-even requirement of the limiting resource, R^* (Fig. 2b–d, ref. 15). The optimal allocation

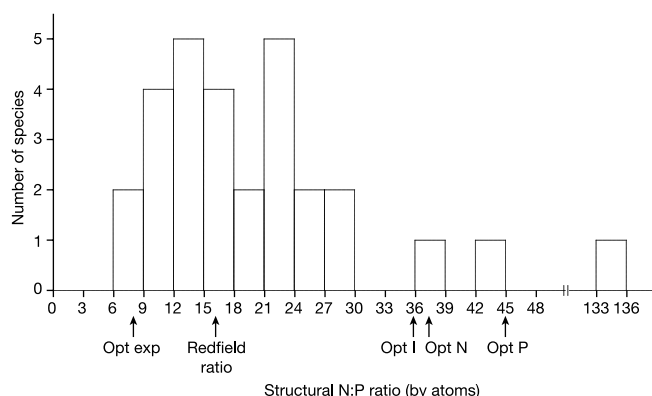


Figure 1 Structural N:P ratio of 29 species of freshwater and marine phytoplankton. The Redfield ratio is shown, as is the theoretical range predicted by the model under the extreme ecological conditions of exponential growth (Opt exp) and competitive equilibrium with light, N and P limiting (Opt I, Opt N and Opt P) conditions.

strategy depends on which resource is limiting: (1) under light limitation, the optimal allocation-to-assembly is $R_a = 0.0447$, giving a structural N:P ratio of 35.8; (2) under N limitation, the optimal strategy is $R_a = 0.0388$, giving a structural N:P ratio of 37.4; and (3) under P limitation, the optimal strategy is $R_a = 0.0164$, giving a structural N:P ratio of 45.0. Colimitation results in an optimal strategy between these values that depends on the resource supply ratio. These optimal N:P ratios roughly bracket the range of variation observed in different species (Fig. 1). The optimal strategy at equilibrium depends on the mortality rate the phytoplankton experience; so conditions of high mortality, such as during intense grazing, lead to greater allocation to assembly-machinery and therefore a lower N:P ratio. In all cases, the optimal strategy balances the conflicting needs for resource acquisition and cellular assembly. Although our parameterization is necessarily rough, given the available data, our qualitative conclusion that the ecological trade-off between rapid growth and equilibrium competitive ability generates a range of optimal N:P ratios is robust to variation in model parameters.

Nitrogen-fixing species often have a higher N:P stoichiometry than non-fixing species. For example, *Trichodesmium* blooms have N:P ratios ranging from 42 to 125 (ref. 16). A simple explanation for N-fixers' high N:P ratios is that they have an exclusive and inexhaustible N supply. Our model suggests another, less obvious explanation. N fixation is energetically costly, leading to a lower ν_I in our model (less carbon incorporated per unit light energy absorbed; see Methods), as observed for *Trichodesmium*¹⁶. Decreasing ν_I increases allocation to light harvesting and decreases allocation to assembly machinery. Therefore, by this mechanism, N fixers have high N:P ratios because they need more light-harvesting machinery to power N fixation at the expense of P-rich assembly machinery. Because both ribosomes and chloroplasts are N-rich, the high N:P ratio of N fixers is mainly due to a low minimum P quota $Q_{\min,P}$.

In our model, for a given environment a single strategy is optimal and displaces all others. Therefore, our approach does not address Hutchinson's paradox of plankton diversity¹⁷. However,

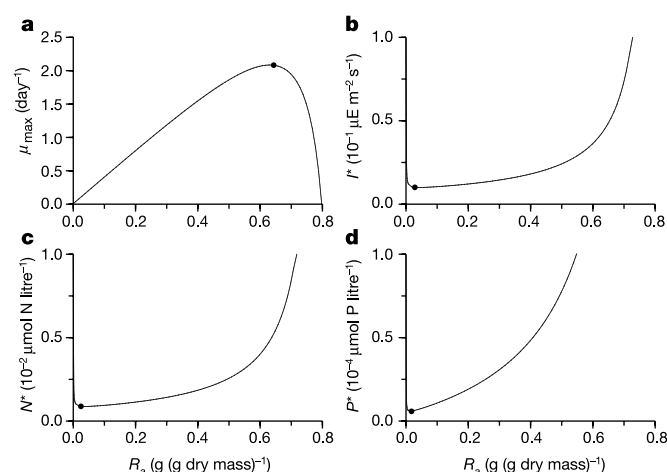


Figure 2 Fitness measures as a function of allocation to assembly machinery. **a**, Maximum growth rate, $\mu_{\max}$. **b–d**, break-even light (**b**), N (**c**) and P (**d**) levels. Species with low R^* values are good equilibrium competitors while those with high $\mu_{\max}$ values are favoured under conditions of exponential growth. The optimal strategies are marked with dots. Note that the allocation to assembly is much higher under exponential growth (**a**) than at competitive equilibrium (**b–d**), and slightly higher under light and N limitation (**b, c**) than under P limitation (**d**). Half-saturation constants used: $K_I = 50 \mu\text{E m}^{-2} \text{s}^{-1}$, $K_N = 5.6 \mu\text{mol N litre}^{-1}$, and $K_P = 0.2 \mu\text{mol P litre}^{-1}$, but these merely scale the graphs in **b–d**.

many mechanisms of coexistence have been explored¹⁸, effectively resolving Hutchinson's paradox. It may be useful to extend our approach to models that permit multispecies coexistence.

The strategic perspective that we take in our model does not specify the level at which variation in allocation occurs. Although species vary in their structural stoichiometries (Fig. 1), intraspecific variation is also possible, owing to either genetic diversity or physiological flexibility. For example, in a series of chemostat growth experiments with a single species, as the dilution rate passed a critical value, the physiological parameters of the culture changed dramatically¹⁹. The 'slow-adapted' cells grown at low-dilution rates had an asymptotic growth rate, μ , of 0.536 day^{-1} and $Q_{\min,P} = 9 \times 10^{-17} \text{ mol P cell}^{-1}$, while the 'fast-adapted' cells had $\mu = 1.13 \text{ day}^{-1}$ and $Q_{\min,P} = 3.61 \times 10^{-16} \text{ mol P cell}^{-1}$. This agrees with our model, which predicts greater allocation to ribosomes and therefore increased P content at high mortality/growth rates.

The Redfield N:P ratio of 16 does not emerge as a universally optimal value from either our empirical survey (Fig. 1) or our theoretical results. Instead, it should be seen as merely the current average stoichiometry of phytoplankton in the ocean, weighted by the relative abundance of species with different structural N:P ratios. In turn, the relative abundance of species is determined by the ecological conditions under which species grow and compete. Competitive equilibrium selects high N:P ratios; exponential growth selects low N:P ratios. Real ecosystems experience a mix of equilibrium and exponential growth phases²⁰, and therefore select species with intermediate N:P ratios (C.A.K., unpublished results). The major determinants of the optimal stoichiometry are the biology and stoichiometry of ribosomes and proteins and the mixture of exponential growth and equilibrium phases; the identity of the limiting resource plays a minor role (Fig. 1). Most of the variation in the optimal N:P ratio in our model results from variation in $Q_{\min,P}$, because all types of machinery contain similar amounts of N.

An implication of this view is that N:P ratios in the ocean could vary over time, simply because of changes in the ecological balance between exponential growth and equilibrium phases, or in N and P availability. Falkowski² asked whether biologists could rule out Broecker and Henderson's hypothesis that N:P ratios rose to 25 during glacial periods²¹. Our results show that this ratio is possible and provide a mechanism that would allow decadal variation in N:P ratios, as has been recently suggested²². Although our model explains structural stoichiometry, the overall stoichiometry (structures plus stores) matches structural stoichiometry during exponential growth⁷, and structural stoichiometry defines the N:P requirements of phytoplankton that diazotrophs match through N-fixation. At equilibrium at low mortality/growth rates, the overall N:P stoichiometry is close to the N:P supply ratio^{6,7}.

We suggest that field surveys²³ should focus less on average values and more on the variation in particulate and dissolved nutrient ratios, by using higher spatial and temporal resolution and presenting the range of values observed rather than just averages. An analysis of long-term time-series data from the North Pacific illustrates the significant temporal variability in N:P stoichiometry²⁴. We await a similarly extensive study of spatial variability in N:P ratios in the sea. Although deep-water stoichiometry integrates over long-time and whole-basin scales and should be relatively constant in time and space, we suspect that greater variability may exist in surface particulate ratios than has been thought, as has been recently documented for C:N ratios²⁵.

We have focused on understanding N:P ratios because N and P are major macroelements in phytoplankton, possibly limiting their growth, and playing an important part in coupled biogeochemical cycles. A recent study has extended the concept of the Redfield-ratio to include trace elements²⁶. Different cellular components may be involved, but we believe that our general modelling approach

linking cellular allocation patterns with competitive ability under different ecological conditions could yield insights into the basis of these other elemental ratios. This approach could also be extended to other ecological trade-offs by considering the different types of cellular machinery. □

Methods

Model and analysis

We use a three-resource version of Droop's model for phytoplankton growth, with equations for cell density, B , and the resource (R) to cell quotas, Q_R . The quota includes both structural and storage pools. Quotas increase with nutrient uptake or photosynthesis, modelled using a Michaelis-Menten function, $f_R(R) = v_R R / (K_R + R)$, where K is the half-saturation constant, and decrease with growth, which is taken to be the minimum of three Droop functions^{7,19}. The growth rate at infinite quota is μ , and $Q_{\min,R}$ is the minimum quota at which growth ceases, corresponding to the amount of this element in the cell's structural machinery. Cell density increases with growth and decreases with density-independent mortality.

$$\begin{aligned} \frac{dB}{dt} &= f_P(P) - \mu \min\left(1 - \frac{Q_{\min,P}}{Q_P}, 1 - \frac{Q_{\min,N}}{Q_N}, 1 - \frac{Q_{\min,C}}{Q_C}\right) Q_P \\ \frac{dQ_N}{dt} &= f_N(N) - \mu \min\left(1 - \frac{Q_{\min,P}}{Q_P}, 1 - \frac{Q_{\min,N}}{Q_N}, 1 - \frac{Q_{\min,C}}{Q_C}\right) Q_N \\ \frac{dQ_C}{dt} &= f_I(I) - \mu \min\left(1 - \frac{Q_{\min,P}}{Q_P}, 1 - \frac{Q_{\min,N}}{Q_N}, 1 - \frac{Q_{\min,C}}{Q_C}\right) Q_C \\ \frac{dB}{dt} &= \mu \min\left(1 - \frac{Q_{\min,P}}{Q_P}, 1 - \frac{Q_{\min,N}}{Q_N}, 1 - \frac{Q_{\min,C}}{Q_C}\right) B - mB \end{aligned} \quad (1)$$

where P , N and I are the available P, N and light levels and m is the mortality rate. These equations are typically supplemented with equations describing the resource dynamics, but because our results do not depend on the exact form of the resource equations, we omit them. Each species follows a set of equations as in (1), and species compete solely through shared use of the resources.

As in general resource-consumer models¹⁵, a species' equilibrium competitive ability for each resource is summarized by the break-even resource level, R^* . For this model:

$$\begin{aligned} P^* &= \frac{m\mu K_P Q_{\min,P}}{v_P(\mu - m) - m\mu Q_{\min,P}} \\ N^* &= \frac{m\mu K_N Q_{\min,N}}{v_N(\mu - m) - m\mu Q_{\min,N}} \\ I^* &= \frac{m\mu K_I Q_{\min,C}}{v_I(\mu - m) - m\mu Q_{\min,C}} \end{aligned} \quad (2)$$

(P^* , N^* , I^*) is the corner of an L-shaped zero-net-growth-isocline (ZNGI) typical of essential resources¹⁵. Phytoplankton can grow when $P > P^*$, $N > N^*$ and $I > I^*$, that is, when resource levels are above the ZNGI, and at equilibrium draw the resources down to some point on the ZNGI. The maximum growth rate realised when a population grows exponentially in a resource-saturated habitat is:

$$\mu_{\max} = \min\left(\frac{\mu v_P}{\mu Q_{\min,P} + v_P}, \frac{\mu v_N}{\mu Q_{\min,N} + v_N}, \frac{\mu v_I}{\mu Q_{\min,C} + v_I}\right) \quad (3)$$

assuming that the quota equilibrates quickly⁷.

We characterize a species by its allocation to four different kinds of cellular machinery: assembly machinery (ribosomes) and three types of resource-acquisition machinery (N- and P-uptake proteins and chloroplasts). These make up R_a , R_N , R_P and R_I proportions of dry mass, respectively. Each type of machinery has its own chemical composition, with N_x the proportion of N and P_x the proportion of P. We assume that these types of machinery constitute a fixed proportion p of the cell's dry mass, which provides a trade-off between the different components.

$$R_a + R_N + R_P + R_I = p \quad (4)$$

The $1 - p$ proportion of biomass that we do not actively consider is assigned its own stoichiometry, (N_o , P_o). The allocation strategy determines the species' maximum assembly rate and resource uptake rates:

$$\mu = \mu' R_a, v_N = v'_N R_N, v_P = v'_P R_P, v_I = v'_I R_I \quad (5)$$

where the primed parameters quantify the efficiency of each kind of machinery. The allocation strategy also determines the N and P content in structural material:

$$\begin{aligned} Q_{\min,N} &= w(R_a N_a + R_N N_N + R_P N_P + R_I N_I + (1 - p)N_o) \\ Q_{\min,P} &= w(R_a P_a + R_N P_N + R_P P_P + R_I P_I + (1 - p)P_o) \end{aligned} \quad (6)$$

where w is cell weight (g dry mass cell⁻¹). The structural N:P ratio is $Q_{\min,N}/Q_{\min,P}$. Together, equations (4)–(6) map the physiological allocation strategy into ecological parameters and structural stoichiometry.

Substituting equations (4)–(6) into equation (3), we find $\mu_{\max}$ as a function of R_a (Fig. 2a). The optimal strategy during exponential growth maximizes $\mu_{\max}$. Substituting equations (4)–(6) into equation (2), we get a formula for R^* for each resource as a function of allocation to assembly machinery (Fig. 2b–d). We find the optimal strategy at equilibrium by minimizing R^* for each resource. Algebraic expressions for the optimal strategies are unwieldy, but can easily be found numerically.

Parameterization

The proportion of cell dry mass allocated to assembly and uptake machinery is set at $p = 0.8$, from the upper-end of the range (0.33–0.8) of protein + RNA (ref. 12). Chemical composition: assembly machinery (ribosomes), $N_a = 0.162$, $P_a = 0.053$; nutrient-uptake

machinery (proteins), $N_N = N_P = 0.17$, $P_N = P_P = 0$; light-harvesting machinery (chloroplasts), $N_I = 0.113$ and $P_I = 0.0032$ (ref. 6). As a neutral choice, we set the composition of the other biomass to the Redfield ratio, $N_o = 0.0631$, $P_o = 0.00873$; our results are relatively insensitive to these values. Assembly-machinery efficiency was set to $\mu' = 4.0 \text{ day}^{-1}$ ($\text{g (g dry mass)}^{-1} \text{ day}^{-1}$) (ref. 13). Maximum carbon-uptake efficiency was set to $v'_I = w$ ($16.8 \text{ g C day}^{-1} \text{ g}^{-1}$), using $P_{\text{opt}}^b = 20 \text{ mg Chr}^{-1}$ (mg chl a^{-1}) (ref. 27), where superscript b indicates normalization to biomass, and the proportion of chlorophyll a in chloroplasts being 0.035 (ref. 28). Maximum N- and P-uptake efficiencies were set to $v'_N = w$ ($3.0 \times 10^3 \text{ g N day}^{-1} \text{ (g N-uptake protein)}^{-1}$) and $v'_P = w$ ($6.7 \times 10^3 \text{ g P day}^{-1} \text{ (g P-uptake protein)}^{-1}$), based on nutrient-transporter turnover times of 0.01 s (refs 29, 30). The minimum carbon quota was set to $Q_{\text{min,C}} = 0.24w$ (ref. 12). The results are independent of cell weight because w cancels from the expression for optimal R_a . The mortality rate was set to $m = 0.01 \text{ day}^{-1}$, chosen from the low end of the observed range to illustrate the lowest feasible value of allocation to assembly machinery. We fix the ratio of uptake machinery types to achieve colimitation during exponential growth, by equating the three terms in the minimum of equation (3). This results in $R_N = 0.00317 R_w$, $R_P = 0.000381 R_w$, and $R_I = 0.996 R_w$, which agrees with the observation that nutrient-uptake proteins are usually a much smaller component of biomass than chloroplasts.

Received 31 December 2003; accepted 27 February 2004; doi:10.1038/nature02454.

1. Redfield, A. C. The biological control of chemical factors in the environment. *Am. Sci.* **46**, 205–221 (1958).
2. Falkowski, P. G. Rationalizing elemental ratios in unicellular algae. *J. Phycol.* **36**, 3–6 (2000).
3. Tyrrell, T. The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**, 525–531 (1999).
4. Lenton, T. M. & Watson, A. J. Redfield revisited 1. Regulation of nitrate, phosphate, and oxygen in the ocean. *Glob. Biogeochem. Cycles* **14**, 225–248 (2001).
5. Rhee, G.-Y. Effects of N:P atomic ratios and nitrate limitation on algal growth, cell composition and nitrate uptake. *Limnol. Oceanogr.* **23**, 10–25 (1978).
6. Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton Univ. Press, Princeton, 2002).
7. Klausmeier, C. A., Litchman, E. & Levin, S. A. Phytoplankton growth and stoichiometry under multiple nutrient limitation. *Limnol. Oceanogr.* (in the press).
8. Rhee, G.-Y. & Gotham, I. J. Optimum N:P ratios and coexistence of planktonic algae. *J. Phycol.* **16**, 486–489 (1980).
9. Bertilsson, S., Berglund, O., Karl, D. M. & Chisholm, S. W. Elemental composition of marine *Prochlorococcus* and *Synechococcus*: implications for the ecological stoichiometry of the sea. *Limnol. Oceanogr.* **48**, 1721–1731 (2003).
10. Heldal, M., Scanlan, D. J., Norland, S., Thingstad, F. & Mann, N. H. Elemental composition of single cells of various strains of marine *Prochlorococcus* *Synechococcus* using X-ray microanalysis. *Limnol. Oceanogr.* **47**, 1732–1743 (2003).
11. Elser, J. J., Dobberfuhl, D., MacKay, N. A. & Schampel, J. H. Organism size, life history, and N:P stoichiometry: towards a unified view of cellular and ecosystem processes. *BioScience* **46**, 674–684 (1996).
12. Geider, R. J. & LaRoche, J. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur. J. Phycol.* **37**, 1–17 (2002).
13. Shuter, B. J. A model of physiological adaptation in unicellular algae. *J. Theor. Biol.* **78**, 519–552 (1979).
14. Kooijman, S. A. L. M. *Dynamic Energy and Mass Budgets in Biological Systems* 2nd edn (Cambridge Univ. Press, UK, 2000).
15. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, NJ, 1982).
16. Letelier, R. M. & Karl, D. M. Role of *Trichodesmium* spp. in the productivity of the subtropical North Pacific Ocean. *Mar. Ecol. Prog. Ser.* **133**, 263–273 (1996).
17. Hutchinson, G. E. The paradox of the plankton. *Am. Nat.* **95**, 137–145 (1961).
18. Tilman, D. & Pacala, S. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. & Schluter, D.) 13–25 (Univ. Chicago Press, Chicago, 1993).
19. Droop, M. R. The nutrient status of algal cells in continuous culture. *J. Mar. Biol. Assoc. UK* **54**, 825–855 (1974).
20. Litchman, E. & Klausmeier, C. A. Competition of phytoplankton under fluctuating light. *Am. Nat.* **157**, 170–187 (2001).
21. Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO_2 changes. *Paleoceanogr.* **13**, 352–364 (1998).
22. Pahlow, M. & Riebesell, U. Temporal trends in deep ocean Redfield ratios. *Science* **287**, 831–833 (2001).
23. Copin-Montegut, C. & Copin-Montegut, G. Stoichiometry of carbon, nitrogen, and phosphorus in marine particulate matter. *Deep-Sea Res.* **30**, 31–46 (1983).
24. Karl, D. M. *et al.* Ecological nitrogen-to-phosphorus stoichiometry at station ALOHA. *Deep-Sea Res.* **48**, 1529–1566 (2001).
25. Schneider, B., Schlitzer, R., Fischer, G. & Nöthig, E.-M. Depth-dependent elemental compositions of particulate organic matter (POM) in the ocean. *Glob. Biogeochem. Cycles* **17**, 1032 (2003).
26. Quigg, A. *et al.* The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature* **425**, 291–294 (2003).
27. Behrenfeld, M. J. & Falkowski, P. G. A consumer's guide to phytoplankton primary productivity models. *Limnol. Oceanogr.* **42**, 1479–1491 (1997).
28. Kirk, J. T. O. & Tilney-Bassett, R. A. E. *The Plastids: Their Chemistry, Structure, Growth and Inheritance* 6 (Freeman Press, London, 1967).
29. Raven, J. A. Nutrient transport in microalgae. *Ad. Microb. Physiol.* **21**, 147–226 (1980).
30. Aksnes, D. L. & Egge, J. K. A theoretical model for nutrient uptake in phytoplankton. *Mar. Ecol. Prog. Ser.* **70**, 65–72 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Falkowski, I. Loladze, S. Pacala and D. Tilman for comments and discussion. We acknowledge support from the Andrew W. Mellon Foundation and the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A.K. (christopher.klausmeier@biology.gatech.edu).

Food-web interactions govern the resistance of communities after non-random extinctions

Anthony R. Ives & Bradley J. Cardinale

Department of Zoology, UW-Madison, Madison, Wisconsin 53706, USA

Growing concern about how loss of biodiversity will affect ecosystems has stimulated numerous studies^{1–5}. Although most studies have assumed that species go extinct randomly^{6–8}, species often go extinct in order of their sensitivity to a stress that intensifies through time (such as climate change)⁹. Here we show that the consequences of random and ordered extinctions differ. Both depend on food-web interactions that create compensation; that is, the increase of some species when their competitors and/or predators decrease in density due to environmental stress. Compensation makes communities as a whole more resistant to stress by reducing changes in combined species densities. As extinctions progress, the potential for compensation is depleted, and communities become progressively less resistant. For ordered extinctions, however, this depletion is offset and communities retain their resistance, because the surviving species have greater average resistance to the stress. Despite extinctions being ordered, changes in the food web with successive extinctions make it difficult to predict which species will show compensation in the future. This unpredictability argues for 'whole-ecosystem' approaches to biodiversity conservation, as seemingly insignificant species may become important after other species go extinct.

Interactions among species make it difficult to predict how ecological communities will respond to environmental degradation¹⁰, for two reasons. First, the sensitivity of an individual species to environmental degradation depends not only on the direct impact of degradation on that species, but also on the indirect effects on that species caused by changes in densities of other species^{11–13}. For example, environmental degradation may decrease the density of competitors and/or predators of a species, thereby causing a compensatory increase in the density of that species^{14–17}. Second, as species go extinct, links within the food web are severed, changing the pathways through which indirect effects operate. Changes in food-web structure depend on the order in which species go extinct, making it difficult to extrapolate from studies that assume extinctions are random to real communities facing progressively intensifying stress from environmental degradation¹⁸.

To disentangle the effects of species interactions on the ability of communities to tolerate environmental degradation, we used mathematical simulations to compare how communities resist changes in abundance as species go extinct randomly versus going extinct in order of their sensitivity to an environmental stress. We

considered communities with three trophic topologies that span a range of community types: tritrophic communities with plants, herbivores, and predators; monotrophic communities comprising just competitors; and communities with arbitrary topology containing prey and predators, competitors, and mutualists (Methods). For each topology, we constructed 1,000 communities in which interaction strengths were chosen from random distributions under the constraints of the specified topology. The direct effects of the stressor on each species were also selected at random, but constrained so that the stressor had a negative effect on all species by decreasing their population growth rates.

We defined the tolerance of a species i as the change in its equilibrium abundance, $x_i^*(s)$, relative to a small change in the magnitude of the stressor s :

$$\delta_i(N) = \frac{\Delta x_i^*(s)}{\Delta s} = \frac{x_i^*(s + \Delta s) - x_i^*(s)}{\Delta s} \quad (1)$$

Negative (positive) values of $\delta_i(N)$ indicate that species i decreases (increases) in abundance in response to intensification of the stressor. $\delta_i(N)$ incorporates both direct and indirect effects of the stressor, and N signifies the dependence of tolerance on the number of surviving species in a community. To measure the ability of the community as a whole to maintain abundance in the face of environmental degradation, we use the average tolerance $\bar{\delta}(N)$ of species in the community.

We subjected each simulated community to a 2×2 factorial numerical experiment (random versus ordered extinctions $\times$ absence versus presence of species interactions). For ordered extinctions, we removed species in order of their tolerance $\delta_i(N)$ to the environmental stressor (Methods). We do not consider the effects of demographic stochasticity on extinction risk^{19–21}, because our focus here is on extinctions caused by environmental degradation. To remove species interactions, we set all interspecific interactions to zero, but retained the same intraspecific interactions and direct impacts of the stressor on species population growth rates.

To illustrate several basic results, we begin with the case of a single, tritrophic community experiencing ordered species extinctions (Fig. 1). In the absence of interspecific interactions, species tolerances $\delta_i(N)$ are independent of the number of species that go extinct. Consequently, the abundance of each species declines monotonically through time as the stressor intensifies (Fig. 1a), and the average tolerance of surviving species, $\bar{\delta}(N)$, increases as more species go extinct (Fig. 1c). Not surprisingly, the rank order of species extinction is readily predicted from the initial tolerance of each species in the community before any species has gone extinct (rank correlation, $r_a = 1$; Fig. 1e).

In contrast, when there are strong interactions among species, species in the food web exhibit density compensation, in which one or more species increase in density as other species decline (Fig. 1b). With each extinction, the pathways of indirect effects of the stressor acting through food-web interactions are rearranged. This rearrangement causes a 're-shuffling' of species tolerances after an extinction, which can be seen by the fluctuations in the species-specific values of $\delta_i(N)$ as N decreases (Fig. 1d). Even though values of $\delta_i(N)$ are re-shuffled, indirect effects cause positive values of $\delta_i(N)$ to be as frequent as negative values, so the average tolerance $\bar{\delta}(N)$ of species in the community remains constant as species go extinct. Despite this constancy, the re-shuffling of species tolerances means that the rank order of species extinction is not as well predicted by the initial tolerances of the species ($r_a = 0.83$; Fig. 1f).

The 2×2 factorial experiment, applied to the 1,000 simulated communities of each topology, leads to two important points. First, ordered extinctions generally increase $\bar{\delta}(N)$ relative to the case of random extinctions, regardless of whether there are interspecific interactions (compare solid to dashed lines, Fig. 2). Thus, ordered

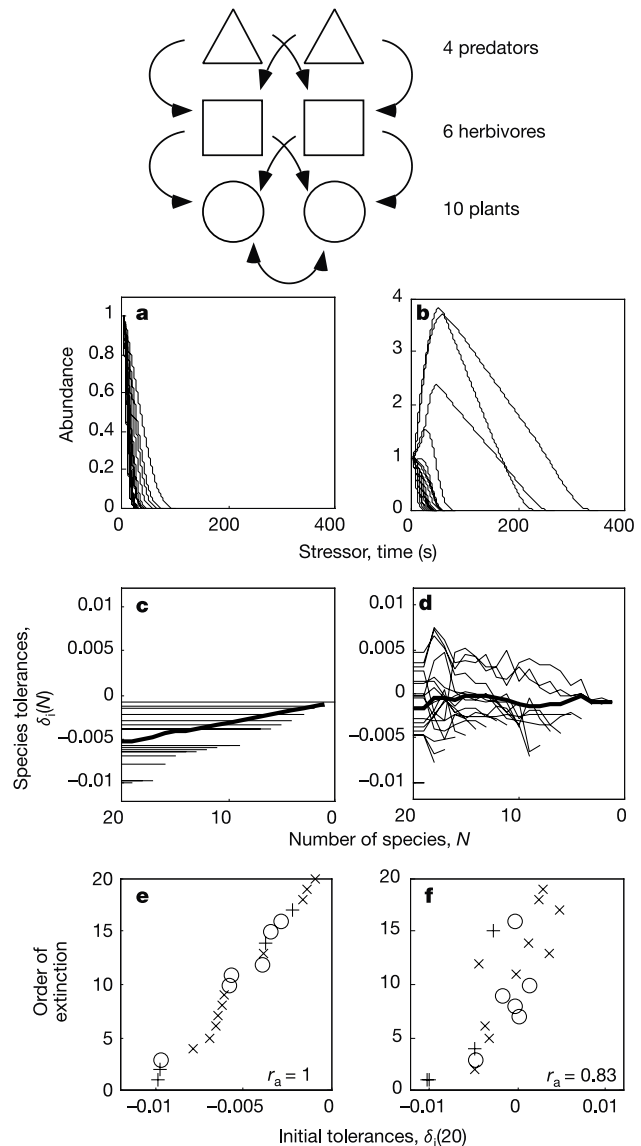


Figure 1 Example of the response of a tritrophic community to environmental stress. **a, b**, Simulations of the dynamics of 20-species communities through time (equation (2)) for food webs with interspecific interactions absent (**a**) or present (**b**). Initial population abundances were scaled to 1, and the stressor was assumed to intensify linearly through time ($s(t) = st$). **c, d**, Species tolerances to the stressor (equation (1)) as a function of the number of species in the communities with interspecific interactions absent (**c**) or present (**d**). Thin lines give $\delta_i(N)$ for individual species, with lines ending at the point of extinction. The heavy line gives the mean tolerance, $\bar{\delta}(N)$, of all surviving species. **e, f**, Rank order of species extinctions versus species tolerances $\delta_i(20)$ in the initial 20-species communities with interspecific interactions absent (**e**) or present (**f**). Plants, herbivores and predators are given by crosses, circles and plus symbols, respectively.

extinctions have a positive effect on the average tolerance of the surviving species. Second, when there are interspecific interactions, the resulting compensation increases $\bar{\delta}(N)$ in the initial 20-species communities (compared to no interactions); $\bar{\delta}(N)$ is zero when $N = 20$ (Fig. 2b–d). However, random extinctions deplete the potential for communities to show density compensation, so $\bar{\delta}(N)$ declines from zero as N decreases. In the presence of interspecific interactions and ordered extinctions (solid lines in Fig. 2b–d), the loss of compensation potential is balanced, almost exactly, by the increase in tolerance associated with the most sensitive species going extinct, leading to little change in $\bar{\delta}(N)$. Finally, food-web inter-

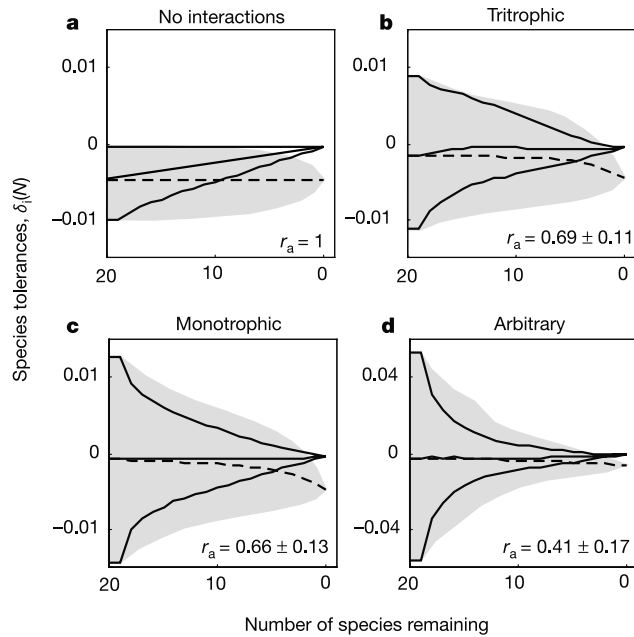


Figure 2 Responses of species tolerances to environmental degradation in various communities. **a**, Communities with interspecific interactions absent; **b**, tritrophic communities; **c**, monotrophic (competitive) communities; and **d**, arbitrarily structured communities. Each panel shows the medians of the minimum, maximum and mean values of $\delta_i(N)$ in each of 1,000 simulations as species went extinct in order of tolerances (solid lines). This procedure was modified by randomly removing species (independently of tolerances), and the resulting medians of minimums and maximums define the shaded region, with the median of the mean values given by the dashed lines. Values for the correlation r_a between the rank order of extinction and species tolerances $\delta_i(20)$ in the initial 20-species communities, for the case of ordered extinctions, are given with their standard deviations.

actions decrease the predictability of extinction, as seen from the reduced correlation between the rank order of initial species tolerances and the rank order of extinction (see r_a in Fig. 2).

To explore the role of interspecific interactions in more detail, we constructed random food webs^{22,23} with 20 species (Methods). We scaled the magnitude of interspecific interactions by p , with $p = 0$ corresponding to no interspecific interactions and $p = 1$ giving interspecific interactions that have on average the same magnitude as intraspecific interactions. Starting with the 20-species communities, we removed the species with the lowest tolerance to the environmental stressor and recomputed the values of $\delta_i(N - 1)$, repeating this procedure until there were only ten surviving species.

Consistent with our previous simulations (Fig. 2), the increase in $\bar{\delta}(N)$ as the most sensitive species go extinct diminishes as the strength of interspecific interactions increases (Fig. 3). This is clear from the change in average tolerance of species, $\Delta\bar{\delta} = \bar{\delta}(10) - \bar{\delta}(20)$, between reduced 10-species communities and the initial 20-species communities. To quantify the re-shuffling of species tolerances caused by extinctions, we computed the correlation r_e between the tolerances of species in the 11- and 10-species communities, $\delta_i(11)$ and $\delta_i(10)$; this correlation drops significantly as p approaches 1. Finally, the impact of interspecific interactions on species tolerances is seen in the correlation r_w between species tolerances in the 10-species community, $\delta_i(10)$, and the values that would have occurred in the absence of interspecific interactions. These results show that even moderately weak interspecific interactions ($p < 0.5$) are sufficient to generate the patterns we have described. The consequences of interspecific interactions can be verified and generalized analytically (Supplementary Information).

We have shown that the consequences of species extinctions for

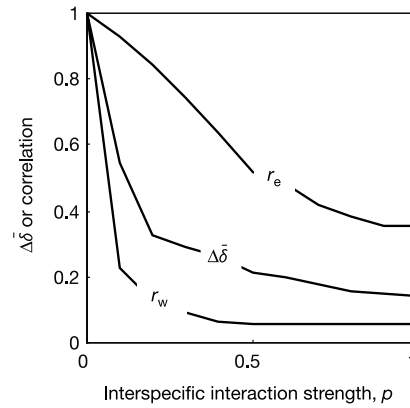


Figure 3 Effect of interspecific interaction strength on the change in mean tolerance of species, $\Delta\bar{\delta}$, when a 20-species community is reduced to a 10-species community through ordered extinction of the most sensitive species. r_w , The correlation between $\delta_i(10)$ for a species and the value of $\delta_i(10)$ that would occur without interspecific interactions. r_e , The correlation between $\delta_i(11)$ and $\delta_i(10)$. To generate estimates, p was increased from 0 to 1 in increments of 0.05, and at each increment 10,000 communities were simulated.

communities depend upon the balance of two opposing forces. First, loss of species from communities (either randomly or non-randomly) tends to decrease the potential for a community to exhibit density compensation that can buffer against future environmental degradation. Second, the ordered extinction of species most sensitive to environmental degradation acts to increase the community-wide average tolerance of the surviving species. When the second of these processes dominates the first, the initial flush of extinctions that often follows the onset of environmental degradation may temper the future impacts of degradation. In contrast, when the first process dominates the second, compensatory dynamics help to alleviate the initial flush of extinctions, thereby allowing the retention of species that may become highly sensitive to the stressor following the initial flush of extinctions. The short-term tolerance of the community to environmental degradation provided by compensation may, therefore, disguise the likelihood of future impacts of degradation. Because density compensation is apparently common and often strong in nature^{4,14,15,24–29}, many communities that have already experienced extinctions may remain unexpectedly sensitive to further environmental degradation.

Species interactions that lead to compensatory dynamics also make it difficult to predict the future sensitivity of a given species to further environmental degradation. This unpredictability argues for whole-ecosystem approaches to ecological conservation, because the future contribution of any particular species to compensation and community resistance is uncertain. Even species that are rare and apparently insignificant may play an important part in community resistance after other species go extinct. □

Methods

Simulations were performed using modified Lotka–Volterra equations:

$$x_i(t+1) = x_i(t)F[(r_i + a_i s(t)) + b_{i,1}x_1(t) + b_{i,2}x_2(t) + \dots + b_{i,N}x_N(t)] \quad (2)$$

where $x_i(t)$ is the abundance of species i at time t , r_i is its intrinsic rate of increase, $b_{i,j}$ is the per capita interaction strength representing the effect of species j on species i (ref. 30), $s(t)$ is the magnitude of an environmental stressor, and a_i governs the direct effect of the stressor on the population growth rate of species i . We assumed that all values of a_i are < 0 , meaning that an increase in the environmental stressor, $s(t)$, leads to a decrease in the per capita population growth rate of every species i . For the discrete-time Lotka–Volterra equations, F is the exponential function, but the results are general for any form of F , provided that species interactions are additive. Although non-additive interactions (either positive or negative) may be common in natural communities, to assume additivity is a reasonable first approximation.

Species tolerances (equation (1)) depend only on values of a_i and $b_{i,j}$, which govern the direct and indirect effects of the stressor (equation (2)); these were selected from uniform random distributions with signs dictated by food-web topology. For tritrophic

communities, values of b_{ij} were selected from uniform distributions between 0 and -0.1 (plant–plant competition), -0.3 (herbivores $\rightarrow$ plants; that is, the effect of herbivores on plants), 0.1 (plants $\rightarrow$ herbivores), -0.1 (predators $\rightarrow$ herbivores), and 0.05 (herbivores $\rightarrow$ predators). Intraspecific interactions were selected at random between -0.1 and -0.2 for plants, and set to -0.2 for herbivores and predators. For competitive communities, interspecific values of b_{ij} were selected uniformly between 0 and -0.1 , and intraspecific values between -0.06 and -0.16 . For the arbitrary topology, the probability of any pair of species interacting was 0.5, and of the interacting pairs of species 45% were competitors, 45% were prey and predators, and 10% were mutualists. The magnitudes of interspecific values of b_{ij} were selected uniformly between 0 and 0.1, with sign dictated by type of interaction, and intraspecific values were selected between -0.06 and -0.16 .

For the random food webs (Fig. 3), we selected intra- and interspecific values of b_{ij} from uniform $(-0.1, 0)$ and $(-0.1, 0.1)$ distributions, respectively. The interaction coefficients b_{ij} were then modified by multiplying interspecific coefficients by p . The vast majority of resulting food webs when $p > 0.5$ were unstable, but comparable analyses constrained to stable food webs gave similar results.

In selecting coefficients, we constrained values for intraspecific interactions b_{ii} to negative numbers. Otherwise, for the case of no interactions (or when communities are reduced to one species) intensifying the stressor increases species abundances. This is because in the absence of species interactions, $\delta_i(N) = -a_i/b_{ii}$, which is positive when $b_{ii} > 0$.

Values of $\delta_i(N)$ were calculated by solving the set of equations satisfied by equation (1) at equilibrium:

$$r_i + a_i s + \sum_j b_{ij} x_j^* (s) = 0 \quad (i = 1, \dots, N) \quad (3)$$

and new values of $\delta_i(N)$ were calculated sequentially as the community size was reduced. Our assumption that the species with the lowest tolerance $\delta_i(N)$ goes extinct first was supported by simulating the full model given by equation (2) (for example, Fig. 1a, b) and categorizing species as extinct once they were reduced in abundance by a factor of 10^{-3} .

Received 5 January; accepted 23 March 2004; doi:10.1038/nature02515.

- Sala, O. E. *et al.* Global biodiversity scenarios for the year 2100. *Science* **287**, 1170–1174 (2000).
- Chapin, F. S. L. *et al.* Consequences of changing biodiversity. *Nature* **405**, 234–242 (2000).
- Ehrlich, P. & Ehrlich, A. *Extinction* (Random House, New York, 1981).
- Terborgh, J. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1926 (2001).
- Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
- Naeem, S. Ecology—Power behind diversity's throne. *Nature* **401**, 653–654 (1999).
- Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
- Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
- Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J. M. Human domination of the Earth's ecosystems. *Science* **277**, 494–499 (1997).
- Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B. & Wood, S. Making mistakes when predicting shifts in species range in response to global warming. *Nature* **391**, 783–786 (1998).
- Ives, A. R. Predicting the response of populations to environmental change. *Ecology* **76**, 926–941 (1995).
- Case, T. J. & Bender, E. A. Testing for higher order interactions. *Am. Nat.* **118**, 920–929 (1981).
- Yodanis, P. *Introduction to Theoretical Ecology* (Harper & Row, New York, 1989).
- Frost, T. M., Carpenter, S. R., Ives, A. R. & Kratz, T. K. In *Linking Species and Ecosystems* (eds Jones, C. G. & Lawton, J. H.) 224–239 (Chapman and Hall, New York, 1995).
- Ives, A. R., Carpenter, S. R. & Dennis, B. Community interaction webs and the response of a zooplankton community to experimental manipulations of planktivory. *Ecology* **80**, 1405–1421 (1999).
- McGrady-Steed, J. & Morin, P. J. Biodiversity, density compensation, and the dynamics of populations and functional groups. *Ecology* **81**, 361–373 (2000).
- Tilman, D. Biodiversity: Population versus ecosystem stability. *Ecology* **77**, 350–363 (1996).
- Ostfeld, R. S. & LoGuidice, K. Community disassembly, biodiversity, loss, and the erosion of an ecosystem function. *Ecology* **84**, 1421–1427 (2003).
- Pimm, S. L., Jones, H. L. & Diamond, J. On the risk of extinction. *Am. Nat.* **132**, 757–785 (1988).
- MacArthur, R. H. *Geographical Ecology* (Harper and Row, New York, 1972).
- Smith, M. D. & Knapp, A. K. Dominant species maintain ecosystem function with non-random species loss. *Ecol. Lett.* **6**, 509–517 (2003).
- May, R. M. *Stability and Complexity in Model Ecosystems* 2nd edn (Princeton Univ. Press, Princeton, New Jersey, 1974).
- Pimm, S. L. *Food Webs* (Chapman and Hall, London, 1982).
- Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
- Peres, C. A. & Dolman, P. M. Density compensation in neotropical primate communities: evidence from 56 hunted and nonhunted Amazonian forests of varying productivity. *Oecologia* **122**, 175–189 (2000).
- Klug, J. L., Fischer, J. M., Ives, A. R. & Dennis, B. Compensatory dynamics in planktonic community responses to pH perturbations. *Ecology* **81**, 387–398 (2000).
- Paine, R. T. Food web complexity and species diversity. *Am. Nat.* **100**, 65–75 (1966).
- Diaz, S., Symstad, A. J., Chapin, F. S., Wardle, D. A. & Huenneke, L. F. Functional diversity revealed by removal experiments. *Trends Ecol. Evol.* **18**, 140–146 (2003).
- Wootton, J. T. Prediction in complex communities: Analysis of empirically derived Markov models. *Ecology* **82**, 580–598 (2001).
- Laska, M. S. & Wootton, J. T. Theoretical concepts and empirical approaches to measuring interaction strength. *Ecology* **79**, 461–476 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Boughman, S. R. Carpenter, A. E. Forbes, R. Haygood, C. T. Harvey, M. R. Helmus, K. J. Tilmon and M. J. Vander Zanden for help. Funding was provided by the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.R.I. (arives@wisc.edu).

Resonance effects indicate a radical-pair mechanism for avian magnetic compass

Thorsten Ritz¹, Peter Thalau², John B. Phillips³, Roswitha Wiltschko² & Wolfgang Wiltschko²

¹Department of Physics and Astronomy, University of California, Irvine, California 92697-4575, USA

²Zoologisches Institut, Fachbereich Biologie und Informatik, J.W. Goethe-Universität, Siesmayerstrasse 70, D-60054 Frankfurt am Main, Germany

³Department of Biology, 2119 Derring Hall, Virginia Tech, Blacksburg, Virginia 24061, USA

Migratory birds are known to use the geomagnetic field as a source of compass information^{1,2}. There are two competing hypotheses for the primary process underlying the avian magnetic compass, one involving magnetite^{3–5}, the other a magnetically sensitive chemical reaction^{6–8}. Here we show that oscillating magnetic fields disrupt the magnetic orientation behaviour of migratory birds. Robins were disoriented when exposed to a vertically aligned broadband (0.1–10 MHz) or a single-frequency (7-MHz) field in addition to the geomagnetic field. Moreover, in the 7-MHz oscillating field, this effect depended on the angle between the oscillating and the geomagnetic fields. The birds exhibited seasonally appropriate migratory orientation when the oscillating field was parallel to the geomagnetic field, but were disoriented when it was presented at a 24° or 48° angle. These results are consistent with a resonance effect on singlet–triplet transitions and suggest a magnetic compass based on a radical-pair mechanism^{7,8}.

The magnetic compass of birds is light-dependent^{9,10}, and exhibits strong lateralization with input coming primarily from the right eye¹¹. However, the primary biophysical process underlying this compass remains unexplained. Magnetite^{3–5,12} as well as biochemical radical-pair reactions^{7,8} have been hypothesized to mediate sensitivity to Earth-strength magnetic fields through fundamentally different physical mechanisms. In the magnetite-based mechanism, magnetic fields exert mechanical forces³. In the radical-pair mechanism, the magnetic field alters the dynamics of transitions between spin states, after the creation of a radical pair through a light-induced electron transfer. These transitions in turn affect reaction rates and products^{7,8}. Although in most radical-pair reactions the effects of Earth-strength magnetic fields are masked by stochastic fluctuations, model calculations¹³ show that such effects can be amplified beyond the level of stochastic fluctuations in specialized radical-pair receptor systems.

Exploiting the principles of magnetic resonance, we developed a diagnostic tool to identify a radical-pair process as the primary process for a physiological magnetic compass. No change in magnetic alignment of magnetite receptors is expected for weak oscillating fields with frequencies larger than 100 kHz (ref. 14).

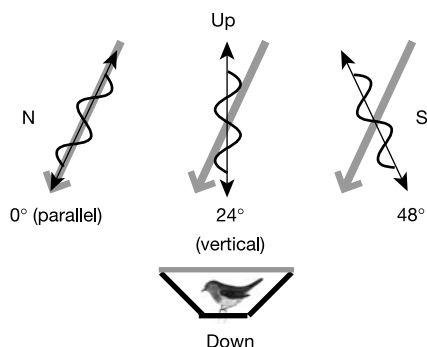


Figure 1 Experimental set-up. Orientation of the 7.0-MHz oscillating magnetic fields (black arrows with sine curve) parallel, at a 24° (vertical) and at a 48° angle to the geomagnetic field (grey arrows; see Fig. 2c–e for results). In the parallel and 48° conditions, the oscillating fields have the same angle with respect to the birds in our experimental set-up.

However, an oscillating magnetic field that is in resonance with the splitting between radical-pair spin states can perturb a radical-pair mechanism by directly driving singlet–triplet transitions. In typical biomolecules, many hyperfine splittings occur in the range of 0.1–10 MHz and a limited number may exist in the range of 10–25 MHz (ref. 15).

We used the migratory orientation of European robins, *Erithacus rubecula*, to detect the possible effects of oscillating magnetic fields on the underlying magnetoreception mechanism. Orientation tests were performed during spring migration under 565 nm light; conditions under which robins normally show excellent orientation using their inclination compass^{16,17}. All birds were tested indoors, in the local geomagnetic field of 46 μ T and 66° inclination. In addition to the control condition (geomagnetic field alone, no oscillating field), we used four experimental conditions in which an oscillating magnetic field was added to the geomagnetic field (Fig. 1).

In the control condition, the robins exhibited seasonally appropriate northerly orientation (Fig. 2a), but in the presence of broadband (0.1–10 MHz, 0.085 μ T) and single-frequency (7 MHz, 0.47 μ T) oscillating fields, both vertically aligned (see Fig. 1), the birds were disoriented (Fig. 2b, d).

To confirm that the observed behavioural change was caused by a direct effect of the oscillating fields on the magnetic compass and not by nonspecific effects due to changes in motivation and so on, we varied the alignment of the 7.0-MHz field. The frequencies at which an oscillating field perturbs a radical-pair reaction depend not only on the chemical nature of the radical pair, but also on the alignment of the oscillating field with respect to the static field¹⁸. This implies that the responses of a magnetic compass system based on radical pairs in the presence of a weak, single-frequency oscillating field can depend on the alignment of the oscillating field, whereas nonspecific effects should occur independently of alignment. We tilted the oscillating field 24° to the north or 24° to the south, so that the two oscillating fields were aligned at the same angle relative to the vertical, but at different angles, parallel and 48°, relative to the geomagnetic field (Fig. 1).

When the oscillating field was parallel to the geomagnetic field, the birds oriented in the migratory direction (Fig. 2c) and their response was indistinguishable from that of the control condition (Table 1). In contrast, when the same oscillating field was presented at 24° and 48° relative to the geomagnetic field, the birds were disoriented (Fig. 2d, e) and their response differed significantly from that of the control condition ($P < 0.01$). The intra-individual scatter in the distribution of nightly headings, as reflected by the length of the birds' mean vectors (r_b), was indistinguishable from that of the control condition when the 7-MHz oscillating field was parallel to the geomagnetic field, but was significantly greater (lower

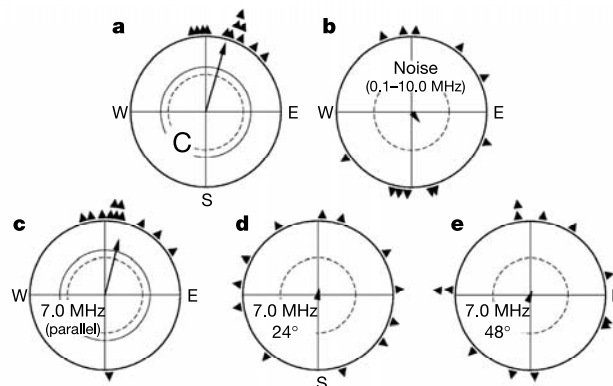


Figure 2 Effects of oscillating magnetic fields on magnetic orientation behaviour of European robins. Triangles indicate the mean headings of the 12 test birds, arrows represent the grand mean vectors (unit length = outer circle radius; see Table 1 for numerical values). The inner circles mark the 5% (dotted) and the 1% significance border of the Rayleigh test²⁷. **a**, Tests in the geomagnetic field only. **b**, Tests in the geomagnetic field and a broadband (0.1–10 MHz) noise field of 0.085 μ T. **c–e**, Tests in a 7.0-MHz field of 0.47 μ T, oriented either parallel (**c**), at a 24° angle (**d**), or at a 48° angle (**e**) to the geomagnetic field.

r_b) in the other three oscillating-field conditions (that is, broadband and 7 MHz at 24° and 48° angles) (see Table 1).

Our findings show that it is unlikely that oscillating fields have an effect on magnetite-based receptors^{3–5,12}, because the dampening effects of the cellular environment prevent magnetite particles from tracking weak radio-frequency magnetic fields. Even in very-low-viscosity physiological conditions (spherical single-domain magnetite in water) we can estimate, following ref. 14, that a 7-MHz field would require an intensity of 285 μ T to produce a noticeable change in alignment, which is far stronger than the 0.47 μ T fields used in our experiments. Likewise, frequencies used in these experiments of less than 10 MHz are far from the expected ferromagnetic resonance frequencies in the GHz range¹⁹, thus rendering thermal or lattice vibration effects of the oscillating fields on magnetite unlikely.

In contrast, resonance effects of oscillating magnetic fields in the frequency range of 0.1–10 MHz are expected in a radical-pair mechanism because hyperfine splittings occur in this range¹⁵. Resonance effects in this frequency range would also be expected in the context of Leask's optical pumping hypothesis⁶, although the lack of evidence for a biological source of energy in the radio-frequency range required by the optical-pumping process⁶ makes this mechanism unlikely.

By what physical mechanism could the remarkably weak oscillating fields used in our experiments (0.085 μ T, 0.47 μ T) affect a radical-pair reaction, and in turn, a radical-pair-based compass system? A simple model calculation (see Methods) suggests that at least in some radical-pair reactions (radical pairs with one dominant hyperfine interaction and a long lifetime), a resonant oscillating magnetic field of a thousandth of the geomagnetic field strength can produce a detectable effect. This remarkable sensitivity to very weak resonant oscillating fields is a noteworthy feature and further studies should analyse the limits of sensitivity in more realistic descriptions of radical pairs.

Our data, together with the above analysis, indicate that a magnetically sensitive radical-pair process in European robins is linked to the physiology of magnetic compass orientation. The most straightforward explanation for our findings is that the radical-pair process indicated by our experiments works as the primary process underlying magnetic compass orientation in European robins and probably in other birds¹⁰. Of course, we cannot exclude the possibility that a radical-pair reaction that is part of an unrelated biochemical pathway was affected. However, the fact that resonance

Table 1 Orientation of European robins in different oscillating magnetic field conditions

Bird	Geomagnetic field only		Noise (0.1–10 MHz)		7.0 MHz parallel		7.0 MHz 24°		7.0 MHz 48°	
	α_b	r_b	α_b	r_b	α_b	r_b	α_b	r_b	α_b	r_b
R 1	26°	0.98	339°	0.24	358°	1.00	110°	0.53	274°	0.45
R 2	20°	0.76	183°	0.42	4°	0.95	126°	0.48	9°	0.84
R 3	47°	0.91	194°	0.61	344°	0.70	86°	0.32	226°	0.98
R 4	350°	0.72	3°	0.21	10°	0.99	17°	0.37	32°	0.20
R 5	15°	0.94	189°	0.74	12°	0.99	162°	0.37	112°	0.85
R 6	1°	0.94	37°	0.90	27°	0.80	330°	0.29	351°	0.17
R 7	18°	1.00	64°	0.42	350°	0.29	297°	0.44	193°	0.73
R 8	20°	0.99	112°	0.51	57°	0.45	220°	0.96	109°	0.11
R 9	354°	0.97	354°	0.80	177°	0.27	58°	0.89	177°	0.32
R 10	24°	0.82	166°	0.09	8°	0.99	261°	0.42	352°	0.15
R 11	358°	0.81	163°	0.84	6°	0.99	278°	0.28	75°	0.80
R 12	37°	0.79	235°	0.47	41°	0.86	3°	0.78	273°	0.31
Grand mean vector	16°, 0.96***		142°, 0.18 ^{n.s.}		14°, 0.78***		11°, 0.10 ^{n.s.}		22°, 0.07 ^{n.s.}	
Median individual vector length	0.93		0.49		0.90		0.43		0.38	
ΔC	C	C	**	***	n.s.	n.s.	**	***	**	**

The α_b and r_b values are based on three recordings of the bird under the respective condition. The grand mean vector is given with its significance by the Rayleigh test indicated by asterisks, followed by the median individual vector length. The bottom line indicates significant differences from the control data obtained in the geomagnetic field only (see Methods for tests). Significance levels: ***, $P < 0.001$; **, $P < 0.01$; n.s., not significant.

effects are only expected in specialized radical-pair systems that can also detect the geomagnetic field^{7,13}, makes it unlikely that a radical-pair process not associated with magnetoreception was affected. There is currently no evidence supporting the existence of such a magneto-sensitive radical-pair process outside the context of magnetoreception and even if one existed, it is uncertain whether it would affect orientation behaviour. In our study we observed no change in activity between birds in oscillating-field and control conditions; and food intake and the general appearance of the birds was normal, suggesting that their health and motivation were unaffected by the brief 75 min exposure to oscillating magnetic fields. In view of this, our findings probably reflect a direct effect of the oscillating fields on the compass mechanism.

This conclusion does not rule out the possibility that birds possess an additional magnetically sensitive system based on magnetite. Magnetite in the form of single domains and super-paramagnetic crystals embedded in specialized structures has been described in the ethmoid region and in the upper beak of migratory birds and pigeons^{20,21}. However, behavioural evidence^{22–24} as well as electrophysiological recordings^{25,26} suggest that the magnetite discovered is not involved in magnetic compass orientation, but instead forms the basis of a magnetic-intensity sensor, potentially used in a magnetic 'map' sense for determining geographic position.

Our study establishes the use of oscillating magnetic fields as a diagnostic tool that can indicate the involvement of a magneto-sensitive radical-pair reaction in birds. Extending this tool to determine the frequency range in which oscillating fields affect the radical-pair mechanism can reveal the chemical nature of the radical pairs involved. Finally, using oscillating magnetic fields as a diagnostic tool is not specific to birds and should be easily transferable to assays with other animal groups. The threshold intensity at which oscillating-field effects can be observed provides information about the underlying mechanism. Behavioural effects from oscillating fields of similar intensity to those used in the present study would suggest a radical-pair mechanism. The absence of behavioural effects from oscillating fields of intensities greater than 50 μ T would make a radical-pair mechanism unlikely. □

Methods

Test birds

European robins are small passerines that migrate at night. The test birds were mist-netted as trans migrants in early September 2002 in the Botanical Garden near the Zoological

Institute in Frankfurt am Main (50° 08' N, 8° 40' E). The birds were kept indoors in individual cages over the winter on a photoperiod that simulated local conditions until December, but was then reduced to 8:16 h light:dark. In the beginning of January 2003, the photoperiod was increased to 13:11 h light:dark. This induced premature *Zugunruhe* (nocturnal migratory restlessness); the experiments took place between 13 January and 17 February 2003.

Test conditions

The tests took place in wooden huts in the garden of the Zoological Institute within the local geomagnetic field of 46 μ T and 66° inclination. To produce the oscillating fields, we modified a test design developed by J.B.P. for similar tests (J.B.P., unpublished), consisting of a coil antenna (210 cm diameter) mounted on a rotatable wooden frame surrounding the test arena. Oscillating currents from a high frequency (HF) generator (Stanford Research Systems DS 34) were amplified by a HF amplifier (Research AF Model 25 W 1,000) and fed into the coil through a resistance of 51 Ω . The coil consisted of a single winding of coaxial cable (RG62A/U, 93 Ω) with 2 cm of the screening removed opposite the feed. The HF field was measured daily, before each test session, using a spectrum analyser (HP89410A) and a magnetic field probe (Rohde & Schwarz, HZ-11816.2770.0, 3 cm probe). Within the test arena, the inhomogeneity of the field was less than 15%. Variations in field intensities between tests were less than 20% of the average value. The HF generator and amplifier were placed outside the huts in varying positions with respect to the test arena. They were switched on during the majority of control tests, but with the power generator turned to zero; comparisons with control tests without this arrangement revealed no observable effect of this procedure.

Test apparatus and procedure

Testing followed standard procedures¹⁶: birds were tested individually in funnel-shaped PVC cages (35 cm upper diameter; 20 cm high) lined with coated paper (BIC Germany, formerly TippEx); the birds left scratches in the coating as they moved. The cages were covered with an opaque plexiglass cover and placed in PVC cylinders, the top of which consisted of a plastic disk carrying the same green light-emitting diodes as those used in earlier studies^{9,16} (peak frequency at wavelength $\lambda = 565$ nm, with $\lambda/2$ at 533 and 583 nm, respectively). The light passed through two diffusers before reaching the bird with an intensity of 2.1 mW m⁻².

The birds were tested once per day. Tests began when the light went out in the housing cages and lasted about 75 min. Each bird was tested three times in each condition. The three tests were arranged in sets; the set of second and third tests began after the set of first and second tests respectively was completed. Within each set, the tests in the various conditions were performed in a pseudo-random order, with the sequence differing between birds.

Data analysis and statistics

For the data analysis, the coated paper was divided into 24 sectors, and the scratches per sector were counted by experimenters that were blind to the test condition. The heading of the respective test was determined by vector addition. From the three headings per test condition for each bird, the mean vector with heading α_b and length r_b was calculated. The twelve α_b values were combined to a grand mean vector, which was tested for directional significance using the Rayleigh test²⁷. The distributions of the birds' α_b values in different conditions were compared using the Mardia-Watson-Wheeler test²⁷. The r_b values, representing the intra-individual variance in locating the migratory direction, are not normally distributed; and so medians are given for each test condition. The r_b values were compared with those obtained under the control conditions using the Wilcoxon test for matched pairs of data.

Model calculations

We used a one-proton radical-pair model²⁸ with an isotropic hyperfine coupling, a , of 0.5 mT, an anisotropy, α of 0.3, and a lifetime of 20 μ s (corresponding to the observed lifetime of flavin-tryptophan radical pairs¹⁵). We solved the stochastic Liouville equation to determine the triplet yield in the presence of a static magnetic field of 46 μ T. We then calculated, by direct numerical integration of the stochastic Liouville equation, the change in triplet yield, $\Delta\Phi_{\text{OMF}}$, caused by an additional 1.3 MHz oscillating magnetic field in resonance with the splitting due to the 46 μ T static field. For comparison, we also calculated the triplet yield change, $\Delta\Phi_{\text{static}}$, resulting from a decrease of 12 μ T in static field, noting that such a change led to disorientation in the magnetic compass orientation responses of robins²⁹. The intensity of the oscillating field required for $\Delta\Phi_{\text{OMF}}$ to equal $\Delta\Phi_{\text{static}}$ is 0.033 μ T, that is, less than any of the intensities employed in our experiments.

Received 2 September 2003; accepted 30 March 2004; doi:10.1038/nature02534.

- Wiltchko, W. & Wiltchko, R. Magnetic compass of European robins. *Science* **176**, 62–64 (1972).
- Wiltchko, R. & Wiltchko, W. *Magnetic Orientation in Animals* (Springer, Berlin, 1995).
- Kirschvink, J. & Gould, J. Biogenic magnetite as a basis for magnetic field detection in animals. *BioSystems* **13**, 181–201 (1981).
- Edmonds, D. T. A sensitive optically detected magnetic compass for animals. *Proc. R. Soc. Lond. B* **263**, 295–298 (1996).
- Shcherbakov, V. P. & Winklhofer, M. The osmotic magnetometer: a new model for a magnetite-based magnetoreceptor in animals. *Eur. Biophys. J.* **28**, 380–392 (1999).
- Leask, M. J. A physico-chemical mechanism for magnetic field detection by migratory birds and homing pigeons. *Nature* **267**, 144–145 (1977).
- Schulten, K. Magnetic field effects in chemistry and biology. *Festkörperprobleme (Adv. Solid State Phys.)* **22**, 61–83 (1982).
- Ritz, T., Adem, S. & Schulten, K. A photoreceptor-based model for magnetoreception in birds. *Biophys. J.* **78**, 707–718 (2000).
- Wiltchko, W., Munro, U., Ford, H. & Wiltchko, R. Red light disrupts magnetic orientation of migratory birds. *Nature* **364**, 525–527 (1993).
- Wiltchko, W. & Wiltchko, R. Magnetic compass orientation in birds and its physiological basis. *Naturwissenschaften* **89**, 445–452 (2002).
- Wiltchko, W., Traudt, J., Güntürkün, O., Prior, H. & Wiltchko, R. Lateralization of magnetic compass orientation in a migratory bird. *Nature* **419**, 467–470 (2002).
- Kobayashi, A. & Kirschvink, J. in *Electromagnetic Fields: Biological Interactions and Mechanisms* (ed. Blank, M.) 367–394 (American Chemical Society Books, Washington DC, 1995).
- Weaver, J., Vaughn, T. & Astumian, R. Biological sensing of small differences by magnetically sensitive chemical reactions. *Nature* **405**, 707–709 (2000).
- Kirschvink, J. *et al.* in *Sensory Transduction* (eds Corey, D. & Roper, S.) 225–240 (Society of General Physiologists 45th Annu. Symp., Rockefeller Univ. Press, New York, 1992).
- Cintolesi, F., Ritz, T., Kay, C., Timmel, C. & Hore, P. Anisotropic recombination of an immobilized photoinduced radical pair in a 50 μ T magnetic field: a model avian photomagnetoreceptor. *Chem. Phys.* **294**, 384–399 (2003).
- Wiltchko, W. & Wiltchko, R. The effect of yellow and blue light on magnetic compass orientation in European Robins, *Erithacus rubecula*. *J. Comp. Physiol. A* **184**, 295–299 (1999).
- Wiltchko, W., Gesson, M. & Wiltchko, R. Magnetic compass orientation of European robins under 565 nm green light. *Naturwissenschaften* **88**, 387–390 (2001).
- Canfield, J., Belford, R., Debrunner, P. & Schulten, K. A perturbation theory treatment of oscillating magnetic fields in the radical pair mechanism. *Chem. Phys.* **182**, 1–18 (1994).
- Kirschvink, J. Microwave absorption by magnetite: a possible mechanism for coupling nonthermal levels of radiation to biological systems. *Bioelectromagnetics* **17**, 187–194 (1996).
- Beason, R. C. & Nichols, J. E. Magnetic orientation and magnetically sensitive material in a transequatorial migratory bird. *Nature* **309**, 151–153 (1984).
- Fleissner, G. *et al.* Ultrastructural analysis of a putative magnetoreceptor in the beak of homing pigeons. *J. Comp. Neurol.* **458**, 350–360 (2003).
- Beason, R. & Semm, P. Does the ophthalmic nerve carry magnetic navigational information? *J. Exp. Biol.* **199**, 1241–1244 (1996).
- Munro, U., Munro, J. A., Phillips, J. B., Wiltchko, R. & Wiltchko, W. Evidence for a magnetite-based navigational 'map' in birds. *Naturwissenschaften* **84**, 26–28 (1997).
- Wiltchko, W., Munro, U., Ford, H. & Wiltchko, R. Effect of a magnetic pulse on the orientation of silvereyes, *Zosterops l. lateralis*, during spring migration. *J. Exp. Biol.* **201**, 3257–3261 (1998).
- Semm, P. & Beason, R. C. Responses to small magnetic variations by the trigeminal system of the bobolink. *Brain Res. Bull.* **25**, 735–740 (1990).
- Beason, R. C. & Semm, P. in *Acta XX Congr. Int. Ornithol.* (ed. Bell, B. D.) 1813–1819 (New Zealand Ornithol. Congr. Trust Board, Wellington, 1991).
- Batschelet, E. *Circular Statistics in Biology* (Academic, London, 1981).
- Timmel, C., Cintolesi, F., Brocklehurst, B. & Hore, P. Model calculations of magnetic field effects on the recombination reactions of radicals with anisotropic hyperfine interactions. *Chem. Phys. Lett.* **334**, 387–395 (2001).
- Wiltchko, W. in *Animal Migration, Navigation and Homing* (eds Koenig, K. & Keeton, W.) 302–310 (Springer, Berlin, 1978).

Acknowledgements We thank the Deutsche Telekom AG, especially H. Küpper, T. Loppnow and B. Marx for technical assistance, and F. Galera, S. Hilmer, C. Koschella and S. Münzner for their help with conducting the experiments. J.B.P. acknowledges the National Science Foundation for financial support. Our work was supported by the Deutsche Forschungsgemeinschaft (W.W.) and the Fetzer Institute (T.R.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.R. (tritz@uci.edu).

Modelling disease outbreaks in realistic urban social networks

Stephen Eubank¹, Hasan Guclu², V. S. Anil Kumar¹, Madhav V. Marathe¹, Aravind Srinivasan³, Zoltán Toroczkai⁴ & Nan Wang⁵

¹Basic and Applied Simulation Science Group, Los Alamos National Laboratory, MS M997, Los Alamos, New Mexico 87545, USA

²Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute 110 8th Street, Troy, New York 12180-3590, USA

³Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742, USA

⁴Centre for Nonlinear Studies and Complex Systems Group, Los Alamos National Laboratory, MS B258, Los Alamos, New Mexico 87545, USA

⁵Department of Computer Science, University of Maryland, College Park, Maryland 20742, USA

Most mathematical models for the spread of disease use differential equations based on uniform mixing assumptions¹ or *ad hoc* models for the contact process^{2–4}. Here we explore the use of dynamic bipartite graphs to model the physical contact patterns that result from movements of individuals between specific locations. The graphs are generated by large-scale individual-based urban traffic simulations built on actual census, land-use and population-mobility data. We find that the contact network among people is a strongly connected small-world-like⁵ graph with a well-defined scale for the degree distribution. However, the locations graph is scale-free⁶, which allows highly efficient outbreak detection by placing sensors in the hubs of the locations network. Within this large-scale simulation framework, we then analyse the relative merits of several proposed mitigation strategies for smallpox spread. Our results suggest that outbreaks can be contained by a strategy of targeted vaccination combined with early detection without resorting to mass vaccination of a population.

The dense social-contact networks characteristic of urban areas form a perfect fabric for fast, uncontrolled disease propagation. Current explosive trends in urbanization exacerbate the problem: it is estimated that by 2030 more than 60% of the world's population will live in cities⁷. This raises important questions, such as: How can an outbreak be contained before it becomes an epidemic, and what disease surveillance strategies should be implemented? Recent studies¹, under the assumption of homogeneous mixing, make the case for mass vaccination in response to a smallpox outbreak. With different assumptions, it has been shown² that mass vaccination is not required. Policymakers must trade off the risks associated with vaccinating a large population⁸ against the poorly understood risks of losing control of an outbreak. Addressing such specific policy questions⁹ requires a higher-resolution description of disease spread than that offered by the homogeneous-mixing assumption and the differential-equations approach.

Here we present a highly resolved agent-based simulation tool (EpiSims), which combines realistic estimates of population mobility, based on census and land-use data, with parameterized models for simulating the progress of a disease within a host and of transmission between hosts¹⁰. The simulation generates a large-scale, dynamic contact graph that replaces the differential equations of the classic approach. EpiSims is based on the Transportation Analysis and Simulation System (TRANSIMS) developed at Los Alamos National Laboratory, which produces estimates of social networks based on the assumption that the transportation infrastructure constrains people's choices about where and when to perform activities¹¹. TRANSIMS creates a synthetic population endowed with demographics such as age and income, consistent with joint distributions in census data. It then estimates positions

and activities of all travellers on a second-by-second basis. For more information on TRANSIMS and its availability, see Supplementary Information. The resulting social network is the best extant estimate of the physical contact patterns among large groups of people—alternative methodologies are limited to physical contacts among hundreds of people or non-physical contacts (such as e-mail or citations) among large groups.

The case study we present is a model of Portland, Oregon, USA, but the approach is broadly applicable. People, in the course of carrying out their daily activities (such as work, study or shopping), move between several locations, both exposing themselves to infectious agents within these locations and transporting those agents between locations. We represent these processes by a social contact network, which can be represented as a bipartite graph, G_{PL} , as shown by the example in Fig. 1a. For Portland, G_{PL} has about 1.6 million vertices, with a giant component of about 1.5 million people and 180,000 locations. The degree distribution of the people vertices in G_{PL} , that is, the number of people Q_j^{PL} who visited j different locations, is shown in Fig. 2a. It has a sharp peak near the average value of about four different locations, followed by a fast, exponentially decaying tail. The degree distribution for the location vertices in G_{PL} is very different, as shown in Fig. 2b. This is the number of locations M_i^{PL} having i different visitors during the day. The distribution has a power-law tail with an exponent of about -2.8 .

For many infectious diseases, transmission occurs mainly between people who are collocated (simultaneously in the same

location), and spread is due mainly to people's movement. Hence we look at two natural projections of G_{PL} obtained by drawing an edge between all pairs of vertices distance two from each other on the bipartite graph, as illustrated in Fig. 1b, c. The result is two disconnected graphs: G_P , containing only people vertices, and G_L , containing only locations. In G_P , the edges are labelled with the sets of time intervals during which the people were collocated. For simplicity, however, we consider $\hat{G}_P$, a static projection of the time-resolved G_P , obtained by discarding time labels, as shown in Fig. 1d. This is reasonable for diseases such as smallpox, severe acute respiratory syndrome or influenza, in which both the incubation period and duration of infectivity are of the order of several days, much longer than the 24-h approximate periodicity of people's contacts. This assumption introduces a systematic bias into results based on $\hat{G}_P$: the static projection yields a worst-case scenario of how the disease is likely to propagate, because $\hat{G}_P$ is much more connected than G_P . Any control strategy that is effective in such worst-case scenarios will also be effective in the time-resolved case. Furthermore, we have modelled the idea of an effective contact (one in which the transmission of disease is likely to occur) by removing from $\hat{G}_P$ all edges for which the duration of contact was less than some minimum threshold, usually one hour. Even this thresholded version of $\hat{G}_P$ is biased towards more connectivity than G_P . An alternative to thresholding is to weight edges according to duration (and other factors affecting transmission). Figure 2c shows the degree distribution of $\hat{G}_P$ for the Portland network. The other important projection of the bipartite graph is the locations network G_L . If there is at least one person travelling from location l_1 directly to l_2 during the day, the two vertices corresponding to locations l_1 and l_2 are connected by a directed edge in G_L from l_1 to l_2 that indicates whether the person is travelling in or out of the location. As before, we form a static version of the locations network, $\hat{G}_L$, by ignoring the time labels on the edges. The in and out degree distributions for the locations network are superimposed in Fig. 2d (ref. 12). The power-law decay evident there shows that $\hat{G}_L$ is a scale-free network⁶ with an exponent of $\gamma \sim -2.8$. A simple explanation for this empirical observation is based on the capacity

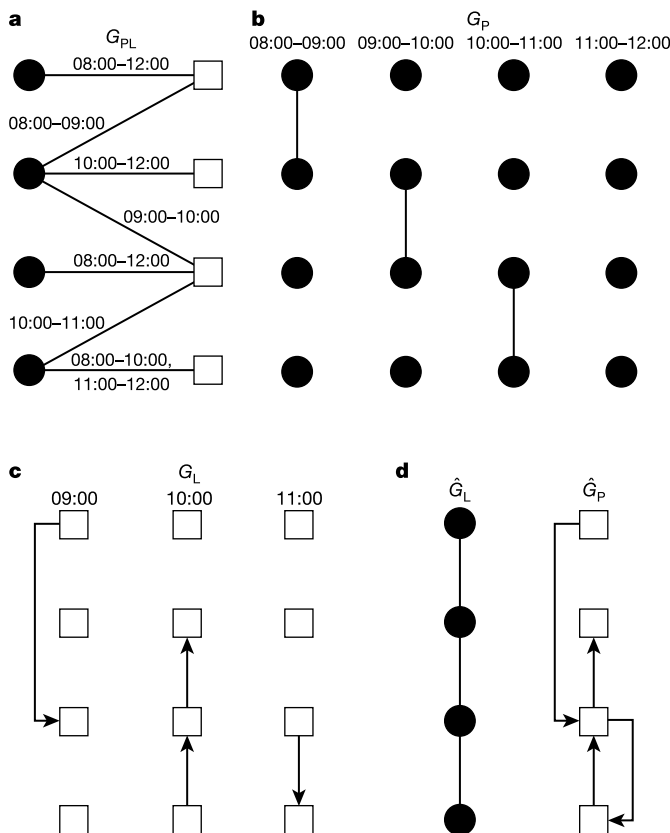


Figure 1 An example of a small social contact network. **a**, A bipartite graph G_{PL} with two types of vertex representing four people (P) and four locations (L). If person p visited location l , there is an edge in this graph between p and l . Vertices are labelled with appropriate demographic or geographic information, edges with arrival and departure times. **b**, **c**, The two disconnected graphs G_P and G_L induced by connecting vertices that were separated by exactly two edges in G_{PL} . **d**, The static projections $\hat{G}_P$ and $\hat{G}_L$ resulting from ignoring time labels in G_P and G_L . People (such as 24-year-old male) are represented by filled circles, and locations (such as 34 Elm Street) by open squares.

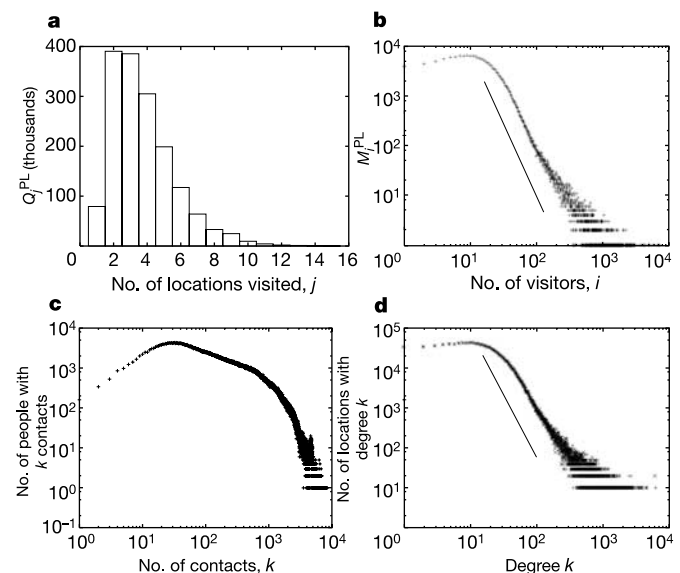


Figure 2 Degree distributions for the estimated Portland social network. **a**, The number of people Q_j^{PL} who visited j different locations in the bipartite people–locations graph G_{PL} . **b**, The number of locations M_i^{PL} in G_{PL} that are visited by exactly i different people. The slope of the straight-line graph is -2.8 . **c**, The number of people who have k neighbours in the static people–contact graph $\hat{G}_P$ on log–log scale. **d**, The in and out degree distributions of the locations network $\hat{G}_L$. The slope of the straight-line graph is -2.8 .

distribution of the locations in Portland. Land-use data indicate that the number of people using a location (its capacity) follows a power-law distribution with the same exponent γ . In large urban areas, people tend to fill locations up to their capacity. More densely filled locations (for example shopping malls) will have a larger number of people moving in from a proportionally larger number of other different locations (for example, homes), which in turn generates the scale-free character of $\hat{G}_L$ with the capacity exponent γ .

Measurements of the average clustering coefficient (see Methods) for $\hat{G}_P$ yield $C_P \approx 0.48$, and for $\hat{G}_L$, $C_L \approx 0.04$, both much larger than the roughly 10^{-6} of an Erdős-Rényi random graph with the same number of vertices and average degree⁵. This, together with the degree distribution and its small diameter (about 6), suggests that the people-contact graph is more like a small-world graph⁵ than a random graph. The clustering coefficient distributions versus degree^{13,14} shown in Supplementary Fig. 1 indicate that the locations network $\hat{G}_L$ is an empirical example of a hierarchical scale-free structure^{15–17}.

Both degree distribution and clustering are relevant to short-term propagation in a network, but longer time dynamics will be driven by global graph properties. It is thus natural to consider estimation schemes for global topological measures, such as expansion (see Methods). Informally, the higher the expansion, the quicker is the spread of any phenomenon (such as disease, gossip or data). We estimated an expansion value of about 2 for G_P by random sampling, indicating that the people-contact graph is extremely connected. An immediate consequence is that, as for an assortatively mixed network¹⁸, $\hat{G}_P$ cannot be shattered by removing (by means of vaccination or quarantine) a small number of high-degree vertices^{19–22}. To verify this, we have computed the size of the giant component—the maximum number of people at risk for disease introduced by a single person—when all vertices of degree more than k are removed. A unique giant component persists even when all vertices of degree 11 and higher are removed, as shown in Fig. 3a. Thus, attempting to shatter the contact graph by vaccinating the most gregarious people in a population would essentially be equivalent to mass vaccination. Similarly, we show in Fig. 3b, c

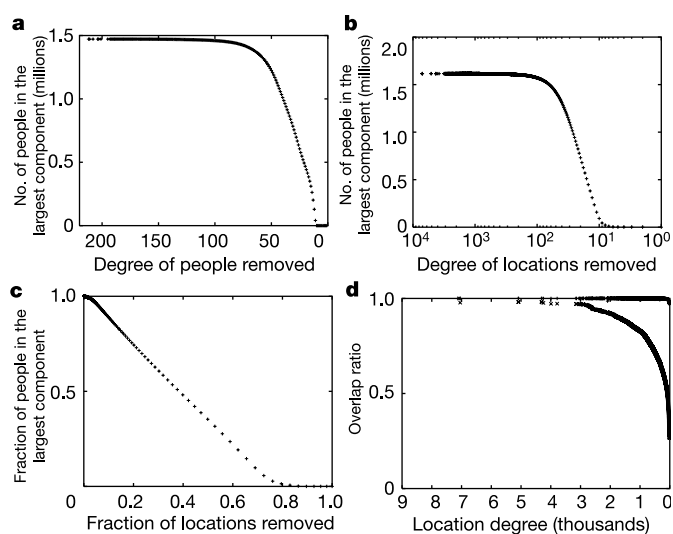


Figure 3 Shattering and covering the people-contact graph. In **a** we remove (by vaccination or quarantine) all people with degree k and higher from the bipartite graph G_{PL} . In **b** and **c** we remove all locations with degree k or higher from G_{PL} and monitor the size of the largest connected component in the static people-contact graph induced by the remaining bipartite graph. **d**, Overlap ratios by degree. The lower curve shows the cumulative overlap ratio by degree, which is the overlap ratio for locations having degree k or less. The upper curve shows the overlap ratio for locations having degree exactly k .

that closing the most-visited locations—or vaccinating everyone who visits them—does not shatter the induced people-contact graph until large fractions of the population have been affected. Other infrastructure networks exhibit very different shattering properties.

Can epidemics be stopped without resorting to mass vaccination? Alternatives rely on early detection and efficient targeting. Here we introduce the overlap ratio, another non-local property of the graph that is crucial to early detection. Consider an idealized situation in which sensors at a location can detect whether any person there is infected. The feasibility of early detection depends on the number of sensors required to cover the population. This problem is equivalent to finding the minimum dominating set²³. That is, we wish to find a subset L' of locations so that all (or most) people visit some location in L' . The overlap ratio²³ $\omega(J)$ of a set of locations $J \subset L$ is $n(J) / \sum_{l \in J} \deg(l)$, where $n(J)$ is the total number of people visiting any location in J , and $\deg(l)$ is the number of different people visiting location $l \in J$. The smaller the overlap ratio, the larger is the number of different locations in J visited by a single person. The overlap ratio by degree, shown in Fig. 3d, is the overlap ratio for the set J_k of locations having degree k . Clearly, not many people visit more than one high-degree location, which implies that the high-degree location vertices form a near-optimal dominating set. With high probability, early identification could be accomplished by using sensors placed at locations with the highest degree.

Alternatives to mass vaccination involve isolating and/or vaccinating small subsets of individuals to ensure that the disease will spread only locally in the graph. Most such strategies assume that people contacted by infectious people are the best candidates for vaccination or quarantine. However, it might also be possible to identify a good subset of the population to target before an outbreak. G_P is composed of tight-knit communities joined by long-range edges. A model for this structure is given by adding long-range edges to random geometric graphs²⁴. An infectious individual (even one with low degree) who travels can nucleate two independent growth centres. Other long-range travellers near these centres can in turn nucleate an exponentially growing number of growth centres, as demonstrated by the rapid worldwide spread of severe acute respiratory syndrome. Thus, targeting long-distance travellers (say, across town for urban regions) is a crucial component of any response.

Our results on the expansion property indicate that disease is likely to spread quickly if not controlled early enough. However, exactly how the number of casualties depends on response delay and

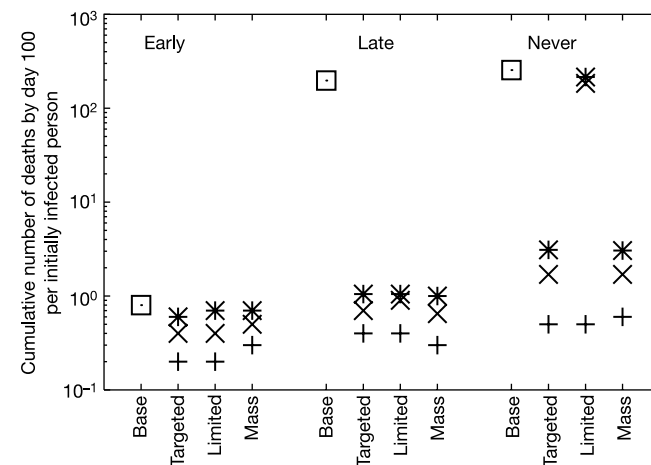


Figure 4 Cumulative number of deaths per number of initial infected, for the case of a smallpox outbreak in downtown Portland, under a number of different response strategies: squares, no vaccine; stars, 10-day delay; multiplication signs, 7-day delay; plus signs, 4-day delay.

what constitutes 'early enough' depend on disease-specific factors such as incubation period and probability of transmission, as well as scenario-specific factors such as the means of introduction. Because these dependences cannot be easily determined from an analysis of the static social network, we turn to simulation, which captures the full time-dependence of G_{PL} (see Supplementary Methods for details).

There is not yet a consensus on models of smallpox. We have designed a model that captures many features on which there is widespread agreement^{9,25} and allows us to vary poorly understood properties through reasonable ranges^{1,2,26}. Our model includes the following features: the incubation period is a gaussian truncated at 7 and 17 days with a 12-day mean and 2-day deviation; the prodromal period is 3–5 days; the infectious period is 4 days, during which infectivity decreases exponentially; death occurs 10–16 days after the rash develops in 30% of normal cases. Ninety-five per cent of susceptibles exposed for 3 h to a person at minimum infectivity will become infected (the remaining 5% have extremely high or low susceptibilities, mimicking some anecdotal transmission incidents); vaccination is assumed to be 100% effective before exposure, and it reduces the mortality and transmission rates when administered up to 2 days after exposure (or 4 days for a previously vaccinated person, assumed to be everyone over the age of 30 years). The model also includes haemorrhagic variants with a shorter incubation period that are 10-fold as infectious and invariably fatal (see Methods and Supplementary Information for details).

Note that EpiSims does not specify a value for R_0 , the basic reproductive number²⁷. This parameter reflects how many people in a susceptible population are directly infected by the introduction of a single infective. R_0 is a convolution of transmission rates and contact patterns, and EpiSims performs the convolution for us. The implied value of R_0 is the ratio of the numbers of people in the first and original cohorts; that is, the number of people initially infected and the number infected directly by them. These estimates obviously include the effects of the simulated response strategy. For the set of experiments reported below, R_0 ranges from 0.4 to 3.4.

In these scenarios, aerosolized smallpox was distributed indoors at busy locations over several hours, infecting of the order of 1,000 people. We assumed that the presence of smallpox was detected on the tenth day after the attack. Furthermore, we assumed that cases could be recognized in the prodromal phase after this date. We did not consider the confounding background distribution of influenza-like symptoms.

We studied the sensitivity of the number of casualties to three factors: mitigation efforts, delay in implementing mitigation efforts, and whether people move about while infectious. We simulated a passive (do nothing) 'baseline' and three active responses: mass vaccination covering 100% of the population in 4 days ('mass'); targeted vaccination and quarantine with unlimited resources ('targeted'); and the same targeted response, using only half as many contact tracers and vaccinators ('limited').

For a movie showing the spatial spread of disease under two different response strategies, see Supplementary Information. Figure 4 compares the efficacy of these strategies. For each strategy we plot (on a logarithmic scale) the ratio of the cumulative number of deaths by day 100 to the number initially infected. The absolute numbers are less important than the rank and relative sizes of gaps between the points. Also shown are the effects of delays of 4, 7 or 10 days in implementing the response. For each of the responses including the baseline, we allowed infected people to isolate themselves by withdrawing to the home. This could be due either to the natural history of the disease, which incapacitates its victims, or to actions taken by public health officials encouraging people to stay home. The results are grouped according to time of withdrawal to the home: (1) early, in which everyone withdraws before becoming infectious, producing the lowest estimates for R_0 ; (2) late, in which everyone withdraws about 24 h after becoming infectious;

and (3) never, in which everyone carries on their daily activities unless they die. The extreme cases are unrealistic but are shown here because they demonstrate the existence of a clear transition.

In this study, time of withdrawal to the home is by far the most important factor, followed by delay in response. This indicates that targeted vaccination is feasible when combined with fast detection. Ironically, the actual strategy used is much less important than either of these factors. □

Methods

Clustering

Here we used the definition of the clustering coefficient, $c_i = 2n_i/[k_i(k_i - 1)]$, given in ref. 5, which measures the extent to which neighbours of a vertex are connected by edges (n_i is the number of connections between the neighbours of vertex i , and k_i is the degree of i). Clustering has important implications for the rate and probability of disease spread^{4,28}.

Graph expansion

The vertex expansion of a set $P' \subset P$ is the ratio between the number of distinct vertices not in P' reached through edges emanating from P' and the number of vertices in P' , denoted by $|P'|$. Clearly, the vertex expansion of a set P' with size $|P'| = \alpha|P|$ is bounded above by $\alpha^{-1} - 1$. By definition, the vertex expansion of the graph G_P is the minimum of the vertex expansions of all sets P' with $|P'| \leq N_P/2$.

Haemorrhagic variants

In our model, haemorrhagic variants occur in 20% of pregnant women, 10% of HIV-positive people, and 2.4% of the population at large. Of those, 30% get an 'early haemorrhagic' variant, with a prodromal period between 0.5 and 1.5 days and an infectious period of 1 day (until death). Pregnancy and HIV status are assigned on the basis of demographics.

Simulation protocol

The protocol we simulated for targeted strategies was to place each prodromal person on a list. Contact tracers chose people at random from the list as they became available. We allowed 24 h for each contact tracer to vaccinate everyone living at the infected person's home and to travel to each location that the infected person visited, vaccinating a fraction of the people there who had been present when the infected person was there. We varied the fraction vaccinated according to the type of activity, from zero at a shopping location to unity at work and home. After 24 h the contact tracer was freed to service the list again. The infected person was sent to a quarantine location. In the 'targeted' case, roughly 20 people were vaccinated for each person initially infected. The peak rate was 10 people per day per initial victim, or roughly 10,000 people per day.

Received 26 November 2003; accepted 5 April 2004; doi:10.1038/nature02541.

- Kaplan, E., Craft, D. & Wein, L. Emergency response to a smallpox attack: the case for mass vaccination. *Proc. Natl Acad. Sci. USA* **99**, 10935–10940 (2002).
- Halloran, M., Longini, I. M. Jr, Nizam, A. & Yang, Y. Containing bioterrorist smallpox. *Science* **298**, 1428–1432 (2002).
- Kretzschmar, M. & Morris, M. Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.* **133**, 165–195 (1996).
- Keeling, M. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B* **266**, 859–867 (1999).
- Watts, D. & Strogatz, S. Collective dynamics of small-world networks. *Nature* **393**, 440–442 (1998).
- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Zwingle, E. Megacities. *Natl Geogr. Mag.* **202**, 70–99 (2002).
- Neff, J. M., Lane, J. M., Fulginiti, V. A. & Henderson, D. A. Contact vaccination: transmission of vaccinia from smallpox vaccination. *J. Am. Med. Assoc.* **288**, 1901–1905 (2002).
- Ferguson, N. et al. Planning for smallpox outbreaks. *Nature* **425**, 681–685 (2003).
- Eubank, S. in *Proc. ACM Symp. Appl. Comput.* (eds Maniatty, W. & Szymanski, B.) 139–145 (ACM Press, New York, 2002).
- Barrett, C. L. et al. *TRANSIMS: Transportation Analysis Simulation System* (Technical Report LA-UR-00-1725, Los Alamos National Laboratory, 2000).
- Chowell, G., Hyman, J. M., Eubank, S. & Castillo-Chavez, C. Scaling laws for the movement of people between locations in a large city. *Phys. Rev. E* **68**, 066102 (2003).
- Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Pseudo-fractal scale-free web. *Phys. Rev. E* **65**, 066122 (2002).
- Szabó, G., Alava, M. & Kertész, J. Structural transitions in scale-free networks. *Phys. Rev. E* **67**, 056102 (2003).
- Jin, E., Girvan, M. & Newman, M. Structure of growing networks. *Phys. Rev. E* **64**, 046132 (2001).
- Ravasz, E., Somera, A., Mongru, D., Olvai, Z. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
- Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
- Newman, M. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65**, 036104 (2002).
- Lloyd, A. & May, R. How viruses spread among computers and people. *Science* **292**, 1316–1317 (2001).
- Callaway, C., Newman, M., Strogatz, S. & Watts, D. Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).

22. Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.* **86**, 3682–3685 (2001).
23. Eubank, S., Anil Kumar, V., Marathe, M. V., Srinivasan, A. & Wang, N. in *Proc. ACM-SIAM Symp. Discrete Algorithms* (ed Munro, I.) 711–720 (SIAM Press, Philadelphia, 2004).
24. Dall, J. & Christensen, M. Random geometric graphs. *Phys. Rev. E* **66**, 016121 (2002).
25. Fenner, F., Henderson, D., Arita, I., Jezek, Z. & Ladnyi, I. *Smallpox and its Eradication* (World Health Organization, Geneva, 1988).
26. Eichner, M. & Dietz, K. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am. J. Epidemiol.* **158**, 110–117 (2003).
27. Keeling, M. & Grenfell, B. T. Individual-based perspectives on R_0 . *J. Theor. Biol.* **203**, 51–61 (2000).
28. Newman, M., Strogatz, S. & Watts, D. Properties of highly clustered networks. *Phys. Rev. E* **68**, 026121 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Korniss, G. Istrate and the Fogarty International Center at the National Institutes of Health for useful discussions, and acknowledge the work of all the members of the TRANSIMS and EpiSims team. The EpiSims project is funded by the National Infrastructure Simulation and Analysis Program (NISAC) at the Department of Homeland Security. The TRANSIMS project was funded by the Department of Transportation. H.G. was supported in part by the National Science Foundation (Division of Materials Research) and Z.T. by the Department of Energy. We thank the anonymous referees for their helpful suggestions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.G.E. (eubank@lanl.gov).

Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus

Christoph Schmidt-Hieber, Peter Jonas & Josef Bischofberger

Physiologisches Institut der Universität Freiburg, D-79104 Freiburg, Germany

Neural stem cells in various regions of the vertebrate brain continuously generate neurons throughout life^{1–4}. In the mammalian hippocampus, a region important for spatial and episodic memory^{5,6}, thousands of new granule cells are produced per day⁷, with the exact number depending on environmental conditions and physical exercise^{1,8}. The survival of these neurons is improved by learning and conversely learning may be promoted by neurogenesis^{8–10}. Although it has been suggested that newly generated neurons may have specific properties to facilitate learning^{2,10,11}, the cellular and synaptic mechanisms of plasticity in these neurons are largely unknown. Here we show that young granule cells in the adult hippocampus differ substantially from mature granule cells in both active and passive membrane properties. In young neurons, T-type Ca^{2+} channels can generate isolated Ca^{2+} spikes and boost fast Na^+ action potentials, contributing to the induction of synaptic plasticity. Associative long-term potentiation can be induced more easily in young neurons than in mature neurons under identical conditions. Thus, newly generated neurons express unique mechanisms to facilitate synaptic plasticity, which may be important for the formation of new memories.

To identify newly generated granule cells in the adult hippocampus, we applied electrophysiological, morphological and immunocytochemical criteria. We made whole-cell recordings from neurons in the granule cell layer and determined the input resistance (R_{in}), which is known to be characteristically different between young and mature neurons¹². Subsequently, we examined the immunoreactivity for polysialic acid neural cell adhesion

molecule (PSA-NCAM), a marker for newly generated granule cells^{1,3,13,14}, and the morphology of the dendritic tree (Fig. 1).

On the basis of the distribution of R_{in} , most cells fell into two categories, with mean R_{in} values of $232 \pm 78 \text{ M}\Omega$ and $4.5 \pm 1.9 \text{ G}\Omega$, respectively (mean $\pm$ s.d.; Fig. 1a, c, e). Cells with high R_{in} were exclusively encountered in the inner granule cell layer adjacent to the subgranular zone, where neural stem cells are located¹. Cells with R_{in} below $400 \text{ M}\Omega$ were immunonegative for PSA-NCAM and showed a complex apical dendritic tree (total dendritic length $3,652 \pm 134 \mu\text{m}$; mean $\pm$ s.e.m.; $n = 14$), indicating that they were mature granule cells. In contrast, all cells with R_{in} larger than $1.5 \text{ G}\Omega$ were immunopositive for PSA-NCAM (Fig. 1d, f; relative fluorescence intensity $76 \pm 6\%$) and had rudimentary apical and basal dendrites (Fig. 1f; total dendritic length $845 \pm 111 \mu\text{m}$, $n = 14$), indicative of immature granule cells^{15–17}. As PSA-NCAM is expressed transiently in granule cells 1–3 weeks after mitosis^{13,14}, these results suggest that the cells with high R_{in} are newly generated immature granule cells in the adult hippocampus. Cells with R_{in} larger than $1.5 \text{ G}\Omega$ were immunonegative for parvalbumin ($n = 4$), suggesting that they do not represent interneurons^{18,19}.

We next studied the active and passive membrane properties of the two classes of neurons (Fig. 2). Both mature and young granule cells showed comparable resting potentials ($-80.8 \pm 0.9 \text{ mV}$ versus $-75.3 \pm 2.0 \text{ mV}$, respectively) and were able to generate fast action potentials with large amplitudes (Fig. 2a–d; $140 \pm 2 \text{ mV}$, $n = 9$ versus $115 \pm 2 \text{ mV}$, $n = 6$). Unlike mature granule cells, which generated trains of action potentials during long current pulses (Fig. 2a; mean action potential frequency $30 \pm 1 \text{ Hz}$), young neurons fired only a few action potentials per stimulus: in 15 of 20 cells, a single action potential was evoked (Fig. 2b). However, the current threshold determined with 100-ms pulses was markedly lower in young neurons than in mature cells ($34 \pm 9 \text{ pA}$, $n = 6$ versus $141 \pm 12 \text{ pA}$, $n = 4$, respectively; $P < 0.01$). In a subset of young neurons currents of less than 10 pA were sufficient to elicit an action potential. Finally, young neurons had a much slower membrane time constant, τ_m , than mature granule cells (Fig. 2e, f; $123 \pm 10 \text{ ms}$, $n = 7$ versus $54 \pm 4 \text{ ms}$, $n = 13$, respectively)²⁰. Thus, newly generated granule cells have membrane properties that favour action potential generation with very small current stimuli, such as the opening of less than ten glutamate receptor channels²¹.

In contrast with mature neurons, young neurons generated transient low-threshold spikes when current injections below the threshold for Na^+ action potential initiation were applied (Figs 2b and 3). Low-threshold spikes were insensitive to $1 \mu\text{M}$ tetrodotoxin (TTX) but were blocked by $50 \mu\text{M}$ Ni^{2+} , showing that they were mediated by T-type Ca^{2+} channels²² (Fig. 3a–c). Ca^{2+} spikes in newly generated neurons had a voltage threshold of $-56.0 \pm 1.2 \text{ mV}$ and an amplitude of $17.6 \pm 1.1 \text{ mV}$ measured from the voltage at the end of the pulse ($n = 16$). Furthermore, they had a slow time course with a duration at half-maximal amplitude of $87 \pm 8 \text{ ms}$ ($n = 16$). Thus, young granule cells generate low-threshold Ca^{2+} spikes under physiological conditions, whereas mature granule cells produce low-threshold Ca^{2+} spikes only in the presence of K^+ -channel blockers²³.

To examine whether the activation of T-type Ca^{2+} channels in young neurons contributed to the initiation of fast Na^+ action potentials, we determined the probability of Na^+ action potential initiation during current pulses of incremental amplitudes (Fig. 3d, e). A concentration of $50 \mu\text{M}$ Ni^{2+} markedly elevated the current threshold for the initiation of fast Na^+ action potentials (Fig. 3e). On average, the current corresponding to the mid-point of the threshold curves increased to $176 \pm 56\%$ of the control value in young neurons ($n = 6$, $P < 0.05$). By contrast, the threshold remained unchanged in mature neurons ($98 \pm 3\%$, $n = 4$). In conclusion, newly generated immature granule cells show a unique

pattern of excitability. As the stimulus intensity is increased, T-type Ca^{2+} channels first generate slow Ca^{2+} spikes and then boost the initiation of fast Na^{+} action potentials.

Postsynaptic action potentials are important associative signals for the induction of spike-timing-dependent plasticity at hippocampal glutamatergic synapses^{24–26}. We therefore tested the impact of postsynaptic excitability on synaptic plasticity in young granule cells. Excitatory postsynaptic potentials (EPSPs) evoked by stimulation of the perforant path were examined in the perforated-patch

configuration (Fig. 4). *In vivo* recordings suggest that dentate gyrus granule cells fire action potentials related to the location of the animal in the spatial environment and the phase of the hippocampal theta oscillation (5–8 Hz)^{27,28}. To mimic this situation, we applied stimulation paradigms corresponding to the physiological firing pattern of a granule cell outside (TBS_0), at the border (TBS_1), or in the centre of its place field (TBS_2 ; Fig. 4a, see Methods). In mature granule cells, TBS_0 and TBS_1 failed to induce long-lasting changes in synaptic strength (Fig. 4b, d, f; the EPSP amplitudes 25–30 min after the end of the induction paradigm were $88 \pm 13\%$ ($n = 5$) and $92 \pm 10\%$ ($n = 6$) of the control value, respectively; $P > 0.4$). However, if bursts of postsynaptic action potentials were evoked during TBS_2 , a long-term potentiation (LTP) of the EPSP amplitude was induced ($214 \pm 46\%$ of control, $n = 6$, $P < 0.05$). Apparently, in mature granule cells a burst of postsynaptic action potentials was required for LTP induction, as reported previously for mature CA1 pyramidal neurons^{25,29}.

Similar to mature granule cells, TBS_0 failed to induce significant changes in EPSP amplitude in young granule cells ($123 \pm 19\%$,

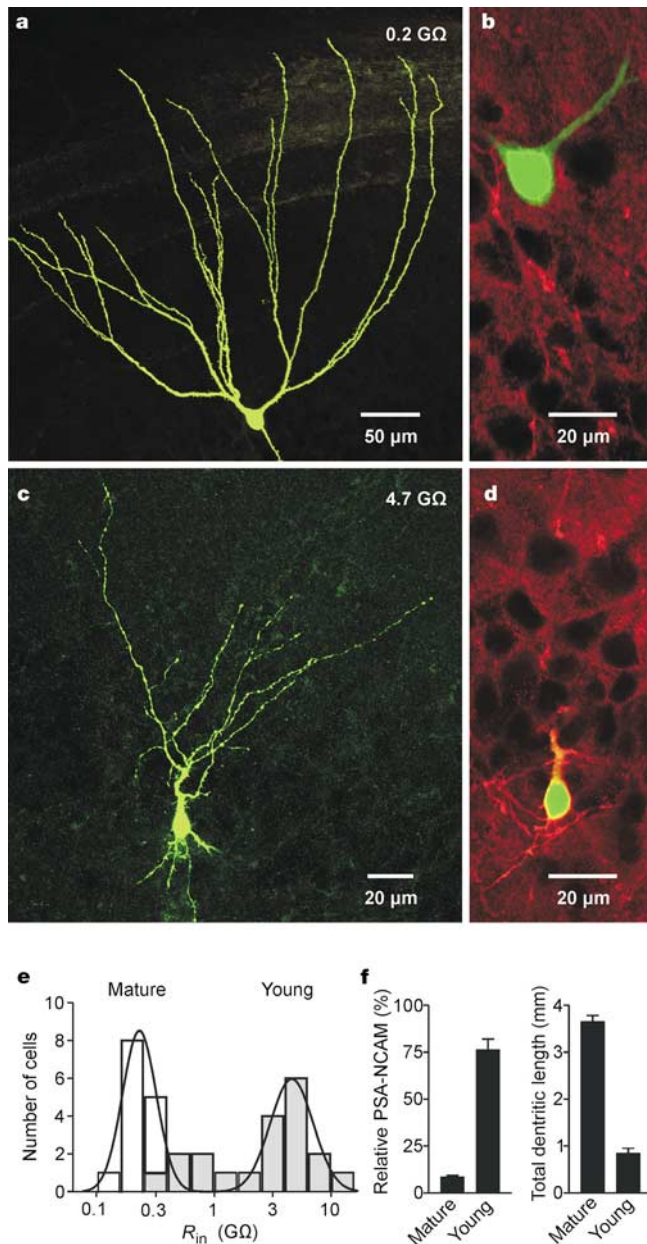


Figure 1 Newly generated granule cells in adult hippocampus are defined by high input resistance, PSA-NCAM immunoreactivity and immature dendritic morphology.

a, c, Biocytin-filled mature (**a**) and young (**c**) granule cell. **b, d**, Confocal images of PSA-NCAM immunoreactivity (red) and biocytin staining (green) of the cells in **a** and **c**, respectively. Note the high PSA-NCAM expression of the immature cell (red dendrite and yellow rim at the soma). **e**, Distribution of input resistance of cells in the inner (grey bars) and outer part of the granule cell layer (open bars). Continuous curves represent the sum of two gaussian functions fitted to the histogram. **f**, Summary graph of PSA-NCAM immunoreactivity and total dendritic length.

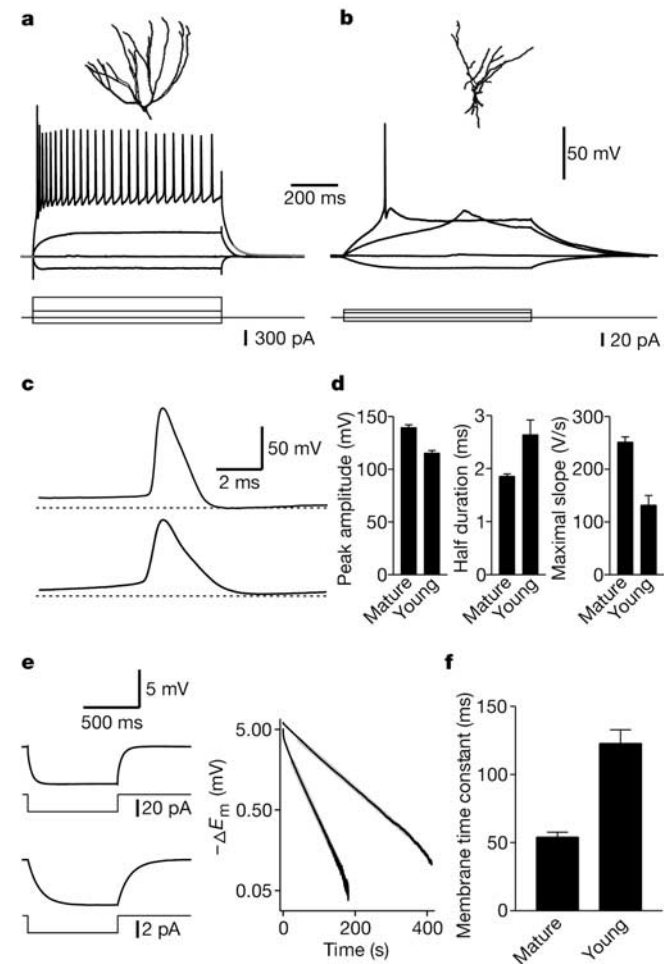


Figure 2 Young granule cells have distinct active and passive membrane properties.

a, Train of action potentials in a mature neuron. **b**, Isolated low-threshold spike and single action potential in a young neuron. **c**, Single action potentials at higher time resolution in a mature (top trace) and a young neuron (bottom trace); dashed lines at -45 mV. **d**, Comparison of peak amplitude, half duration and maximal slope of action potentials. **e**, Voltage changes evoked by hyperpolarizing current pulses in a mature (top trace) and a young neuron (bottom trace). Right panel shows logarithmic plot of voltage changes (ΔE_m) after the pulse; regression lines (grey) are shown superimposed. **f**, Summary graph of membrane time constants.

$n = 5, P > 0.1$). However, unlike in mature cells TBS₁ was sufficient to induce LTP (Fig. 4c, e, g; the EPSP amplitude was $154 \pm 9\%$ of control, $n = 9, P < 0.01$). Furthermore, in the young cells TBS₂ had only slightly larger effects than TBS₁ ($162 \pm 10\%$, $n = 5$). Statistical comparison between young and mature cells further indicates that TBS₁ has significantly larger effects in the young neurons ($P < 0.001$). Thus, our results demonstrate that excitatory synapses on newly generated granule cells show associative plasticity, and that the threshold for induction is lower than in mature neurons. Pairing of EPSPs with single postsynaptic action potentials is a highly efficient stimulus for the induction of LTP in young granule cells, whereas a burst of postsynaptic action potentials is required in mature neurons. Hence, different induction rules for LTP apply to young and mature granule cells.

Newly generated granule cells show specialized membrane properties and enhanced associative plasticity at their glutamatergic input synapses, suggesting a causal link. One scenario is that the high input resistance and the expression of T-type Ca^{2+} channels promote LTP because they facilitate the activation of neurons by sparse glutamatergic synaptic inputs. Furthermore, Ca^{2+} and Na^{+} action potentials as evoked by TBS₁ in young granule cells may efficiently induce LTP by generating both a global rise in postsynaptic Ca^{2+} concentration and a maximal relief of Mg^{2+} block of NMDA-type glutamate receptors.

Functional plasticity such as TBS₁-induced LTP may have profound structural consequences. It is well established that PSA-NCAM selectively labels newly generated neurons 1–3 weeks after

mitosis^{13,14}. In this time period the fate of the newly generated neurons (survival versus cell death) is determined³⁰. Enriched environment and spatial learning increase the proportion of surviving neurons^{1,9}, suggesting the possibility that associative LTP promotes the survival and final integration of newly generated granule cells into the adult hippocampal circuit.

Both functional and structural plasticity in newly generated cells may be important for the formation of spatial memory. Single spikes during the theta cycle are observed *in vivo* when the animal approaches the border of the place field of a hippocampal granule cell^{27,28}. In the centre of the place field, however, granule cells fire brief bursts of action potentials. If the induction rules for synaptic plasticity as reported here also apply *in vivo*, our data would suggest that in young granule cells LTP will occur in a much larger 'induction area' as compared with mature neurons. This might be

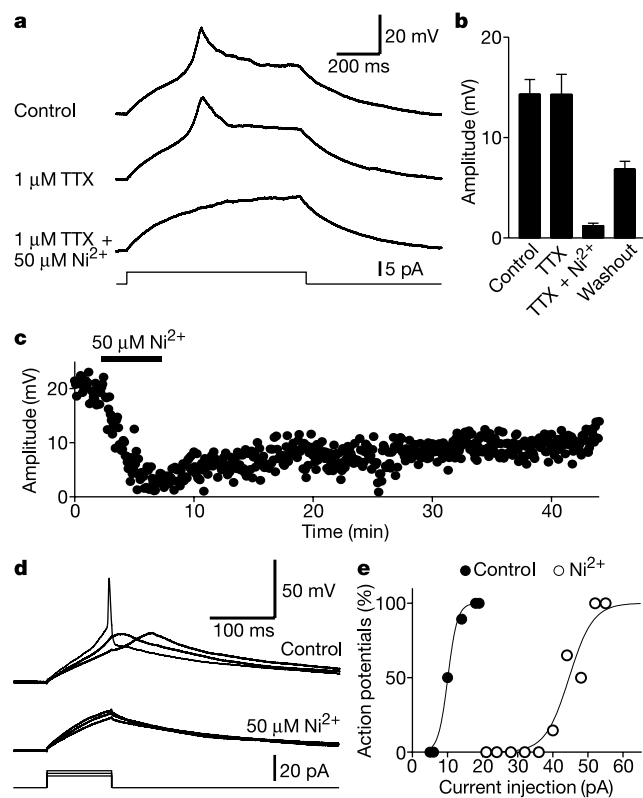


Figure 3 T-Type Ca^{2+} channels lead to enhanced excitability in young granule cells. **a, b**, The low-threshold spike in young neurons was insensitive to TTX, but was blocked by $50 \mu\text{M Ni}^{2+}$, indicating that it is mediated by T-type Ca^{2+} channels. Spike amplitude was measured from the potential at the end of the current injection. **c**, Plot of Ca^{2+} spike amplitude against time in another experiment. The spike was reversibly blocked by Ni^{2+} . **d, e**, The current threshold for the initiation of Na^{+} action potentials was increased when T-type Ca^{2+} channels were blocked by Ni^{2+} . Continuous curves in **e** represent fitted Boltzmann functions.

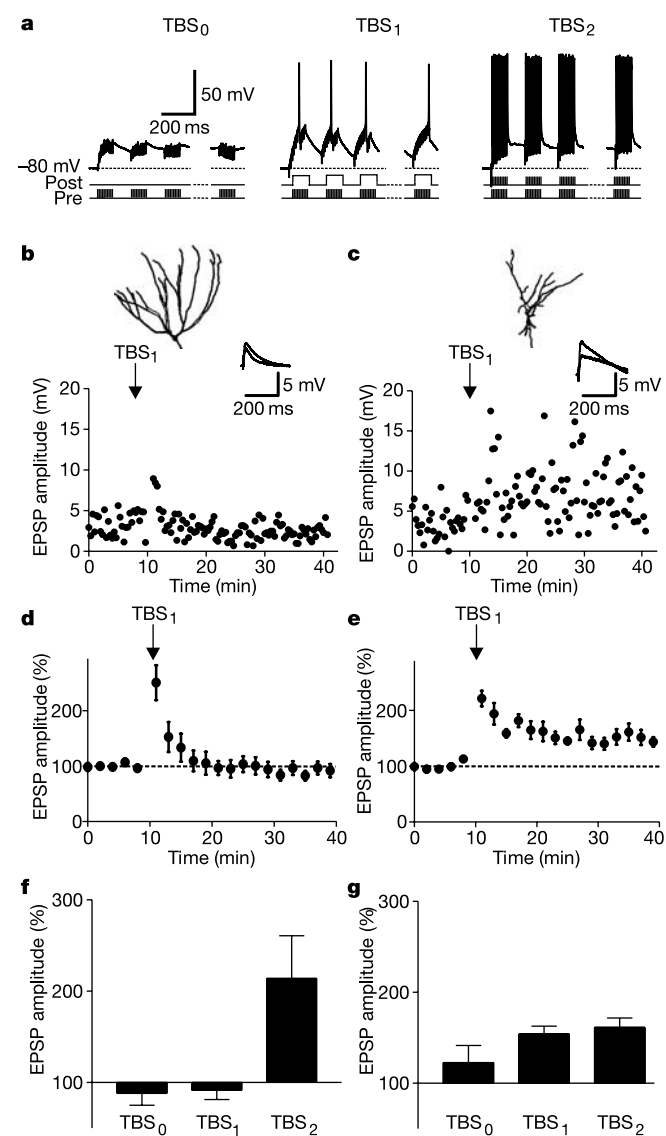


Figure 4 Different induction rules for long-term potentiation in young and mature neurons. **a**, Theta-burst induction paradigms, exemplified here for a young granule cell. **b, c**, Plot of EPSP amplitude against time in a mature (**b**) and a young (**c**) granule cell. **d, e**, Average plots for mature (**d**) and young (**e**) cells. In mature cells, TBS₁ had no effect, whereas it led to significant LTP in young cells. Data points represent means of six consecutive events. **f, g**, Summary of LTP experiments. TBS₁ led to LTP only in young neurons, whereas TBS₂ led to LTP in both mature and young cells. All data were from perforated-patch recordings.

important for the formation of new place fields. As the dendritic tree grows and the cell becomes fully integrated into the hippocampal circuit, a burst of action potentials will be necessary for LTP induction. Thus, the induction area will be much more confined, finally leading to a high spatial selectivity and a precise neuronal encoding of new spatial environment⁶. □

Methods

Slice preparation

Adult male Wistar rats (2–3 months old, approximately 300 g) were reared in large cages with running wheels and climbing frames. The animals were anaesthetized by isoflurane and killed by decapitation, in accordance with institutional guidelines. Transverse 350- μ m-thick slices were cut from the hippocampus using a vibratome (DTK-1000; Dosaka). For the dissection and the storage of the slices, a solution containing 64 mM NaCl, 25 mM NaHCO₃, 10 mM glucose, 120 mM sucrose, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 0.5 mM CaCl₂ and 7 mM MgCl₂ (equilibrated with 95% O₂/5% CO₂) was used. Slices were incubated at 35 °C for 30 min and subsequently held at room temperature.

Immunohistochemistry and morphology

Cells were filled with biocytin (1 mg ml⁻¹) during whole-cell recording and subsequently fixed in 4% paraformaldehyde. After washing, tissue sections were incubated overnight at 4 °C with the primary antibody for PSA-NCAM (mouse, Chemicon, 1:400) together with 5% goat serum and 0.3% Triton X-100. The secondary antibody (goat anti-mouse-Alexa546, 1:200, Molecular Probes) was applied together with fluorescein isothiocyanate-conjugated avidin-D (2 μ l ml⁻¹, Vector) and 0.3% Triton X-100 for 24 h at 4 °C. After washing, slices were embedded in Prolong Antifade (Molecular Probes). Parvalbumin immunoreactivity was tested using an antibody against parvalbumin (mouse, Swant, 1:5,000) and an Alexa568-conjugated secondary antibody. Fluorescence was analysed with a confocal laser-scanning microscope (LSM 510, Zeiss).

PSA-NCAM immunofluorescence was measured in 1- μ m optical sections and quantified as intensity per pixel along the somatic cell membrane. Background intensity was subtracted, and the signal was normalized by the mean intensity of the brightest 1% of pixels within the image (325 $\times$ 325 μ m). The morphology of the cells was reconstructed in three dimensions from a stack of 50–100 confocal images using NeuronTracer software (Bitplane) running on an O2 workstation (Silicon Graphics). The total dendritic length was calculated as the sum of the lengths of all dendritic segments. Only cells with intact dendrites were used.

Electrophysiology

Slices were superfused with a physiological extracellular solution containing 125 mM NaCl, 25 mM NaHCO₃, 25 mM glucose, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 2 mM CaCl₂ and 1 mM MgCl₂ (equilibrated with 95% O₂/5% CO₂). A concentration of 10 μ M bicuculline methiodide was added in plasticity experiments, and 10 μ M bicuculline and 10 μ M 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) in membrane time constant experiments. Recordings were performed under visual control. Patch pipettes (4–9 M Ω) were pulled from thick-walled borosilicate glass tubing. Voltage signals were measured with an Axopatch 200A amplifier (Axon Instruments) in the I-clamp fast mode, filtered at 5 kHz, and digitized at 10 or 20 kHz using a 1401plus interface (Cambridge Electronic Design). For whole-cell recordings, pipettes were filled with a solution containing 130 mM K-gluconate, 20 mM KCl, 2 mM MgCl₂, 4 mM K₂ATP, 0.3 mM NaGTP, 10 mM Na₂-phosphocreatine and 10 mM HEPES (pH 7.2). For perforated-patch recordings, a pipette solution containing 140 mM K-gluconate, 20 mM KCl, 5 mM MgCl₂, 10 mM HEPES, 0.1 mM EGTA, 50–100 μ g ml⁻¹ amphotericin B and 0.2–0.4% DMSO was used. Bridge balance was used to compensate the series resistance of 10–50 M Ω (whole cell) or 60–90 M Ω (perforated patch). Resting potential was determined in a subset of recordings with seal resistances >5 R_{in}. The holding potential was set to -80 mV. To stimulate the medial perforant path, 2–4-M Ω pipettes filled with HEPES-buffered Na⁺-rich solution were used. The pipettes were placed in the centre of the molecular layer close to the dendrites of the recorded cell, and 200- μ s pulses were applied at a frequency of 0.05 Hz. EPSP amplitude measurements were started 10–15 min after a stable access resistance had developed. Recordings were performed at 21–23 °C.

LTP induction protocols

The LTP induction paradigms consisted of brief presynaptic bursts (ten stimuli at 100 Hz) repeated ten times at theta frequency (5 Hz). This presynaptic stimulation was applied alone (TBS₀), paired with a slow postsynaptic depolarization that typically evoked a single action potential (TBS₁), or paired with a train of brief stimuli that evoked a burst of postsynaptic action potentials^{25,29} (TBS₂; 5-ms delay between pre- and postsynaptic train). Paradigms were applied four times at 0.1 Hz. Only one TBS paradigm was tested per cell. If no LTP occurred, a maximal high-frequency stimulation protocol was applied (100 Hz, 1 s, four times, paired to postsynaptic depolarizations to 0 mV), and experiments were rejected if no LTP could be induced (8 of 24 mature and 4 of 23 young cells).

Data analysis

The input resistance was determined from the voltage at the end of 500- or 800-ms current pulses leading to approximately 5 mV hyperpolarization. To determine the membrane time constant, 200–400 voltage traces were averaged, and the logarithmically transformed decay was analysed by linear regression. For the quantification of LTP, 15 or 30 consecutive

EPSPs were averaged 10 min before the onset and 25–30 min after the end of the induction protocol. Peak amplitudes were measured from holding potential and data are given as mean $\pm$ s.e.m. unless specified differently. A Wilcoxon signed rank test or a Mann–Whitney test were used to assess statistical significance.

Received 2 February; accepted 8 April 2004; doi:10.1038/nature02553.

Published online 25 April 2004.

- Gage, F. H. Mammalian neural stem cells. *Science* **287**, 1433–1438 (2000).
- Nottebohm, F. Why are some neurons replaced in adult brain? *J. Neurosci.* **22**, 624–628 (2002).
- Seki, T. & Arai, Y. Highly polysialylated neural cell adhesion molecule (NCAM-H) is expressed by newly generated granule cells in the dentate gyrus of the adult rat. *J. Neurosci.* **13**, 2351–2358 (1993).
- van Praag, H. *et al.* Functional neurogenesis in the adult hippocampus. *Nature* **415**, 1030–1034 (2002).
- Lisman, J. E. Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron* **22**, 233–242 (1999).
- Burgess, N., Maguire, E. A. & O'Keefe, J. The human hippocampus and spatial and episodic memory. *Neuron* **35**, 625–641 (2002).
- Cameron, H. A. & McKay, R. D. G. Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. *J. Comp. Neurol.* **435**, 406–417 (2001).
- van Praag, H., Christie, B. R., Sejnowski, T. J. & Gage, F. H. Running enhances neurogenesis, learning, and long-term potentiation in mice. *Proc. Natl Acad. Sci. USA* **96**, 13427–13431 (1999).
- Gould, E., Beylin, A., Tanapat, P., Reeves, A. & Shors, T. J. Learning enhances adult neurogenesis in the hippocampal formation. *Nature Neurosci.* **2**, 260–265 (1999).
- Shors, T. J. *et al.* Neurogenesis in the adult is involved in the formation of trace memories. *Nature* **410**, 372–376 (2001).
- Snyder, J. S., Kee, N. & Wojtowicz, J. M. Effects of adult neurogenesis on synaptic plasticity in the rat dentate gyrus. *J. Neurophysiol.* **85**, 2423–2431 (2001).
- Liu, Y.-B., Lio, P. A., Pasternak, J. F. & Trommer, B. L. Developmental changes in membrane properties and postsynaptic currents of granule cells in rat dentate gyrus. *J. Neurophysiol.* **76**, 1074–1088 (1996).
- Seki, T. Hippocampal adult neurogenesis occurs in a microenvironment provided by PSA-NCAM-expressing immature neurons. *J. Neurosci. Res.* **69**, 772–783 (2002).
- Seki, T. Expression patterns of immature neuronal markers PSA-NCAM, CRMP-4 and NeuroD in the hippocampus of young adult and aged rodents. *J. Neurosci. Res.* **70**, 327–334 (2002).
- Lübbers, K. & Frotscher, M. Differentiation of granule cells in relation to GABAergic neurons in the rat fascia dentata. Combined Golgi/EM and immunocytochemical studies. *Anat. Embryol.* **178**, 119–127 (1988).
- Rihn, L. L. & Claiborne, B. J. Dendritic growth and regression in rat dentate granule cells during late postnatal development. *Dev. Brain Res.* **54**, 115–124 (1990).
- Rao, M. S. & Shetty, A. K. Efficacy of doublecortin as a marker to analyse the absolute number and dendritic growth of newly generated neurons in the adult dentate gyrus. *Eur. J. Neurosci.* **19**, 234–246 (2004).
- Jonas, P., Bischofberger, J., Fricker, D. & Miles, R. Interneuron diversity series: Fast in, fast out—temporal and spatial signal processing in hippocampal interneurons. *Trends Neurosci.* **27**, 30–40 (2004).
- Liu, S. *et al.* Generation of functional inhibitory neurons in the adult rat hippocampus. *J. Neurosci.* **23**, 732–736 (2003).
- Spruston, N. & Johnston, D. Perforated patch-clamp analysis of the passive membrane properties of three classes of hippocampal neurons. *J. Neurophysiol.* **67**, 508–529 (1992).
- Spruston, N., Jonas, P. & Sakmann, B. Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. *J. Physiol. (Lond.)* **482**, 325–352 (1995).
- Eliot, L. S. & Johnston, D. Multiple components of calcium current in acutely dissociated dentate gyrus granule neurons. *J. Neurophysiol.* **72**, 762–777 (1994).
- Blaxter, T. J., Carlen, P. L. & Niesen, C. Pharmacological and anatomical separation of calcium currents in rat dentate granule neurones *in vitro*. *J. Physiol. (Lond.)* **412**, 93–112 (1989).
- Debanne, D., Gähwiler, B. H. & Thompson, S. M. Cooperative interactions in the induction of long-term potentiation and depression of synaptic excitation between hippocampal CA3–CA1 cell pairs *in vitro*. *Proc. Natl Acad. Sci. USA* **93**, 11225–11230 (1996).
- Magee, J. C. & Johnston, D. A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209–213 (1997).
- Normann, C. *et al.* Associative long-term depression in the hippocampus is dependent on postsynaptic N-type Ca²⁺ channels. *J. Neurosci.* **20**, 8290–8297 (2000).
- Jung, M. W. & McNaughton, B. L. Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus* **3**, 165–182 (1993).
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
- Pike, F. G., Meredith, R. M., Olding, A. W. A. & Paulsen, O. Postsynaptic bursting is essential for 'Hebbian' induction of associative long-term potentiation at excitatory synapses in rat hippocampus. *J. Physiol. (Lond.)* **518**, 571–576 (1999).
- Dayer, A. G., Ford, A. A., Cleaver, K. M., Yassae, M. & Cameron, H. A. Short-term and long-term survival of new neurons in the rat dentate gyrus. *J. Comp. Neurol.* **460**, 563–572 (2003).

Acknowledgements We thank M. Frotscher, E. Förster and G. Stuart for critically reading the manuscript, and K. Winterhalter and A. Blomenkamp for technical assistance. Work was supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B. (josef.bischofberger@uni-freiburg.de).

Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors

Wei-hai He^{1,*}, Frederick J.-P. Miao¹, Daniel C.-H. Lin¹, Ralf T. Schwandner^{1,2}, Zhulun Wang¹, Jinhai Gao^{1,*}, Jin-Long Chen¹, Hui Tian¹ & Lei Ling¹

¹Tularik Inc., 1120 Veterans Boulevard, South San Francisco, California 94080, USA

²Tularik GmbH, Josef-Engert-Strasse 9, 93053 Regensburg, Germany

* Present addresses: Genospectra Inc., 6519 Dumbarton Circle, Fremont, California 94555, USA (W.H.); Novartis Institute of Biomedical Research Inc., 100 Technology Square, Cambridge, Massachusetts 02139, USA (J.G.)

The citric acid cycle is central to the regulation of energy homeostasis and cell metabolism¹. Mutations in enzymes that catalyse steps in the citric acid cycle result in human diseases with various clinical presentations². The intermediates of the citric acid cycle are present at micromolar concentration in blood and are regulated by respiration, metabolism and renal reabsorption/extrusion. Here we show that GPR91 (ref. 3), a previously orphan G-protein-coupled receptor (GPCR), functions as a receptor for the citric acid cycle intermediate succinate. We also report that GPR99 (ref. 4), a close relative of GPR91, responds to α -ketoglutarate, another intermediate in the citric acid cycle. Thus by acting as ligands for GPCRs, succinate and α -ketoglutarate are found to have unexpected signalling functions beyond their traditional roles. Furthermore, we show that succinate increases blood pressure in animals. The succinate-induced hypertensive effect involves the renin-angiotensin system and is abolished in GPR91-deficient mice. Our results indicate a possible role for GPR91 in renovascular hypertension, a disease closely linked to atherosclerosis, diabetes and renal failure^{3,6}.

In a search for natural ligands for orphan GPCRs, we tested extracts from various animal tissues for their ability to evoke an increase in intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) using the aequorin assay⁷. We found that fractions from pig kidney extracts specifically activated cells expressing GPR91 (Fig. 1a). GPR91 is an orphan GPCR highly expressed in the kidney and shares 33% amino acid identity with GPR99/GPR80 (refs 4, 8). On the basis of their homology with the purinergic receptor P2Y1, nucleotide ligands were predicted for GPR91 and GPR99 (ref. 4). However, the GPR91 ligand activity in pig kidney extracts was resistant to various stringent treatments including alkaline phosphatase, peptidase, and hydrolysis in 6 M HCl at 100 °C. Accordingly, the supposition that GPR91 might be activated by a nucleotide or peptide ligand was unlikely. We purified the natural ligand for GPR91 by ion-exchange, size-exclusion and reversed-phase fast performance liquid chromatography/high-performance liquid chromatography (Fig. 1a).

A major molecular ion $[\text{M} + \text{H}]^+$ at m/z 119.2 was observed by mass spectrometry (Fig. 1b). ¹H NMR analysis revealed a single type of proton in the highly purified GPR91 ligand (Fig. 1c). ¹³C NMR analysis further suggested the presence of $-\text{CH}_2-$ (methylene) and $=\text{C}=\text{O}$ (carbonyl) groups (Fig. 1d). Combined with mass spectrometry results and the biochemical properties of the ligand, the purified GPR91 ligand was predicted and confirmed to be succinic acid (Fig. 1c, d).

Commercially obtained succinate (the physiological form of succinic acid) increased $[\text{Ca}^{2+}]_i$ dose-dependently in the aequorin assay (Fig. 2a). Succinate also activated mouse and rat orthologues of GPR91 (Fig. 2a). The succinate-induced increase in $[\text{Ca}^{2+}]_i$ was further confirmed with a fluorimetric imaging plate reader (FLIPR) system in the 293-hGPR91 cell line (293 cells stably expressing human GPR91) (Fig. 2b). The concentration giving half-maximal response (EC_{50}) for succinate-induced activation of human GPR91 was $56 \pm 8 \mu\text{M}$ in the aequorin assay and $28 \pm 5 \mu\text{M}$ in the FLIPR assay. GPR91 was selectively activated by succinate but not by other citric acid cycle intermediates (Supplementary Information) or any of a collection of 800 pharmacologically active compounds and known GPCR ligands tested (data not shown). Furthermore, of 200

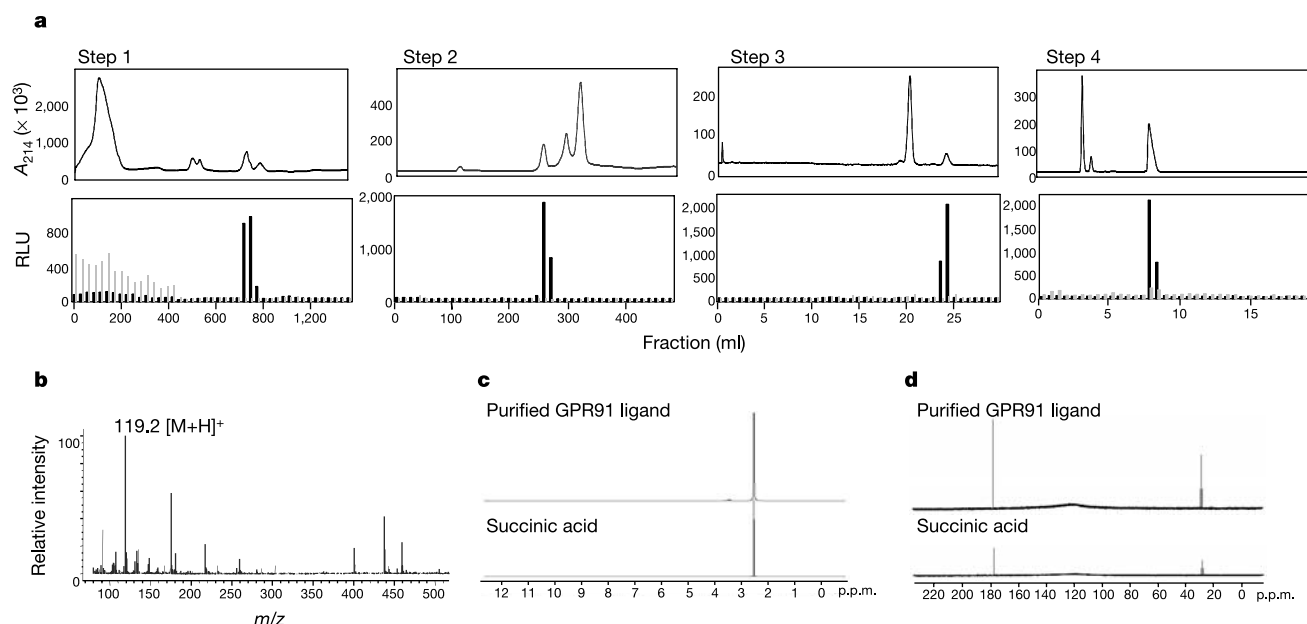


Figure 1 Purification and identification of GPR91 ligand from pig kidney extracts.

a, Purification of GPR91 ligand through four steps of fast performance liquid chromatography/high-performance liquid chromatography (step 1, anion exchange; step 2, size exclusion; step 3, size exclusion; step 4, reversed phase). The activity of each

fraction was assayed by the aequorin assay in CHO cells transfected with GPR91 (black bars) or control plasmid (grey bars). RLU, relative luminescence unit. **b**, Mass spectrometric analysis of the purified GPR91 ligand. **c**, ¹H NMR (**c**) and ¹³C NMR (**d**) of the final purified GPR91 ligand: analysis and comparison with succinic acid.

carboxylic acids and structurally related analogues, succinate was found to be the most potent in increasing $[Ca^{2+}]_i$ (a representative subset is summarized in Supplementary Information).

Because GPR99 is the closest homologue of GPR91, we proposed that other citric acid cycle intermediates and carboxylic acids might activate GPR99. We identified α -ketoglutarate as a ligand for human GPR99, with an EC_{50} of $69 \pm 11 \mu M$ in the aequorin assay (Fig. 2c) and $32 \pm 4 \mu M$ in the FLIPR assay (Fig. 2d). α -Ketoglutarate also activated mouse GPR99 (Fig. 2c). Neither succinate nor α -ketoglutarate activated calcium mobilization with about 30 GPCRs unrelated to GPR91 and GPR99 (data not shown).

To dissect the signalling pathways of GPR91 and GPR99, various biochemical assays were investigated. Succinate inhibited forskolin-stimulated cAMP production in 293-hGPR91 cells, an effect abolished by preincubation with pertussis toxin (PTX) (Fig. 2e). In addition, succinate induced the accumulation of inositol phosphate in 293-hGPR91 cells (Fig. 2f). The succinate-induced accumulation of inositol phosphate and an increase in $[Ca^{2+}]_i$ were each partly inhibited by pertussis toxin (Fig. 2f and Supplementary Information). Moreover, succinate activated extracellular signal-regulated kinase (Erk) in 293-hGPR91 cells (Supplementary Information). Taken together, these results suggest that GPR91 activation by succinate couples to at least two signalling pathways, a pertussis-toxin-sensitive G_i/G_o pathway and a pertussis-toxin-insensitive G_q pathway. Similarly, α -ketoglutarate stimulated inositol phosphate formation in a GPR99-dependent manner (Fig. 2g).

Because GPR99 activation did not affect cAMP levels (data not shown), and both inositol phosphate formation (Fig. 2g) and $[Ca^{2+}]_i$ flux (Supplementary Information) were found to be insensitive to pertussis toxin, GPR99 seems to act exclusively through a G_q -mediated pathway.

Ligand-induced receptor internalization is often characteristic of GPCR activation and signal attenuation⁹. Immunofluorescent staining of cells expressing Flag-tagged GPR91 or GPR99 revealed that both GPR91 and GPR99 were localized to the plasma membrane (Fig. 2h). Ligand stimulation induced the internalization of both receptors (Fig. 2h).

Because the dicarboxylate groups of succinate and α -ketoglutarate are required for the activation of GPR91 and GPR99, we speculated that basic residues of GPR91 and GPR99 might be important in binding dicarboxylate ligands. A partial three-dimensional model of GPR91 was generated. Mutation of Arg 99, His 103, Arg 252 or Arg 281 (Supplementary Information) abolished GPR91 activation by succinate, whereas mutation of His 249, Arg 255 or Tyr 107 had no effect on receptor function (Fig. 3a). Various GPR91 mutants were expressed at similar levels (Fig. 3b) and localized at the plasma membrane (data not shown). All four residues required for GPR91 activation by succinate (Arg 99, His 103, Arg 252 and Arg 281) are directed into the central cavity where *cis*-retinal binds in the rhodopsin structure (Fig. 3c, d). These four positively charged residues cluster together, Arg 99 and His 103 of helix III on one side and Arg 252 of helix VI and Arg 281 of helix VII on the other side,

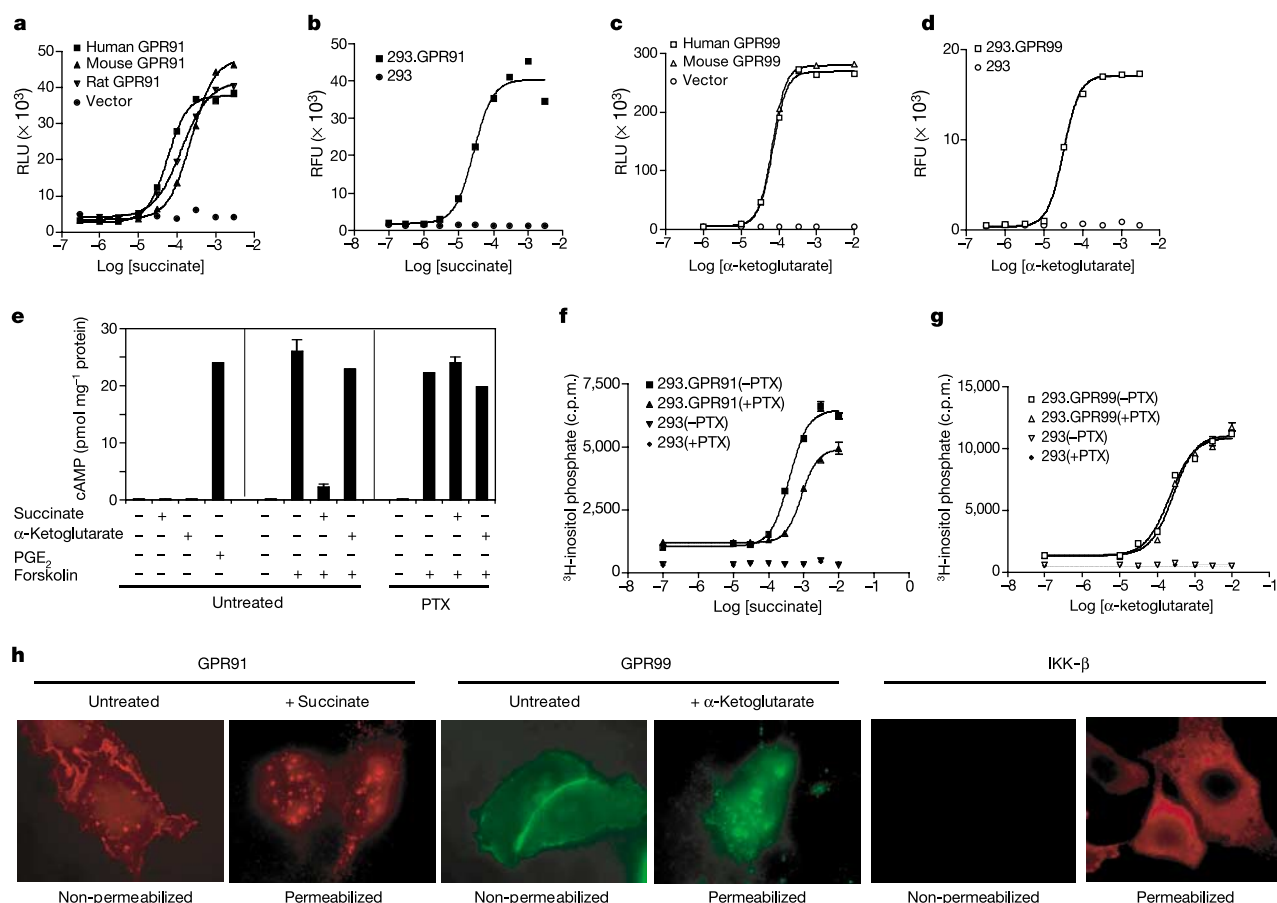


Figure 2 Pharmacological characterization. Data are shown as means $\pm$ s.e.m. for three or four determinations. **a**, GPR91 activation by succinate in aequorin assay. **b**, FLIPR assay in 293-hGPR91 cells. RFU, relative fluorescence unit. **c**, **d**, Activation of GPR99 by α -ketoglutarate in aequorin (**c**) and FLIPR assays (**d**). **e**, Effect of succinate (200 μM) on intracellular cAMP in 293-hGPR91 cells. Pertussis toxin and forskolin were included as indicated. Prostaglandin E₂ (PGE₂) was used as a control. **f**, Inositol phosphate formation

in 293-hGPR91 cells and the effect of pertussis toxin (PTX). **g**, Inositol phosphate accumulation in 293-hGPR99 cells and the effect of PTX. **h**, Immunofluorescence staining. Plasma membrane localization of the receptors (in non-permeabilized cells) and the internalized vesicles (in permeabilized cells) were visualized by anti-Flag antibody. IKK- β , I κ B kinase- β . An intracellular protein IKK- β^{30} served as a control.

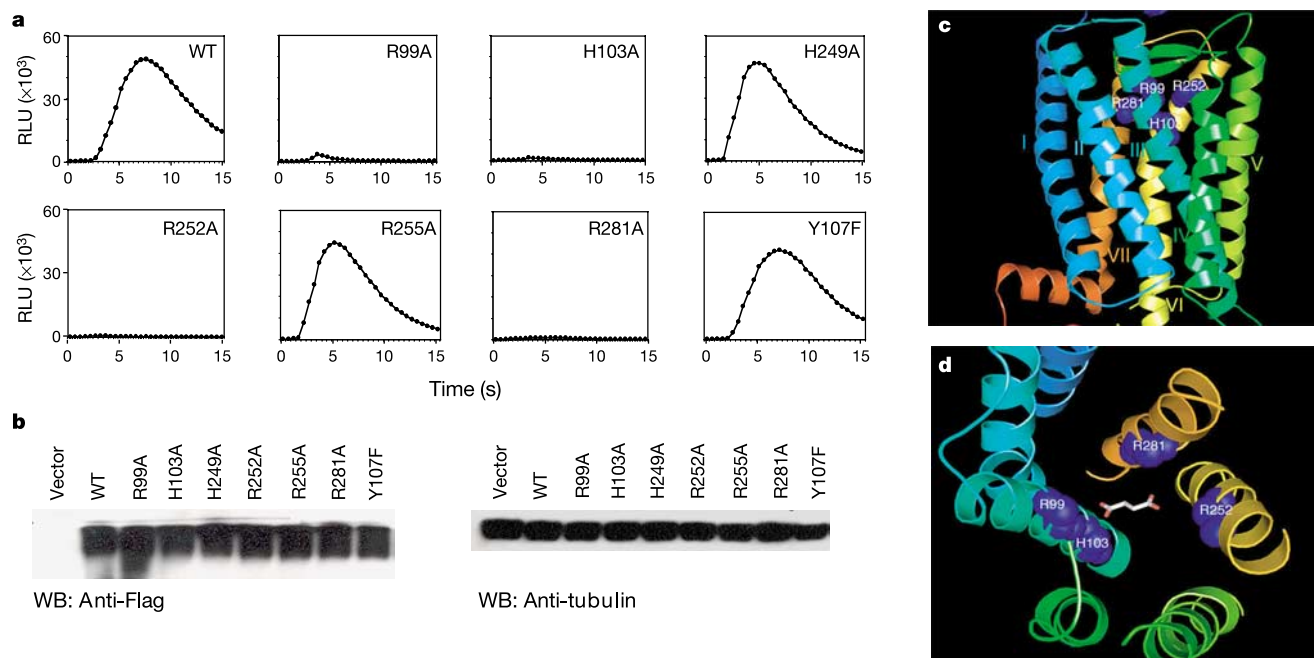


Figure 3 Mutational analysis and predicted three-dimensional model for GPR91. **a**, Activation of wild-type (WT) and mutant GPR91 receptors by succinate (200 μ M) in aequorin assay in CHO cells. **b**, The expression of wild-type and mutant GPR91 receptors in CHO cells as determined by western blot (WB) analysis. **c**, Predicted sites for Arg 99,

His 103, Arg 252 and Arg 281 in the three-dimensional ribbon presentation of GPR91. The main chains of these four residues are coloured in blue. **d**, Top view from the extracellular side of the four basic residues in the ligand-binding pocket with a succinic acid molecule placed in the cavity.

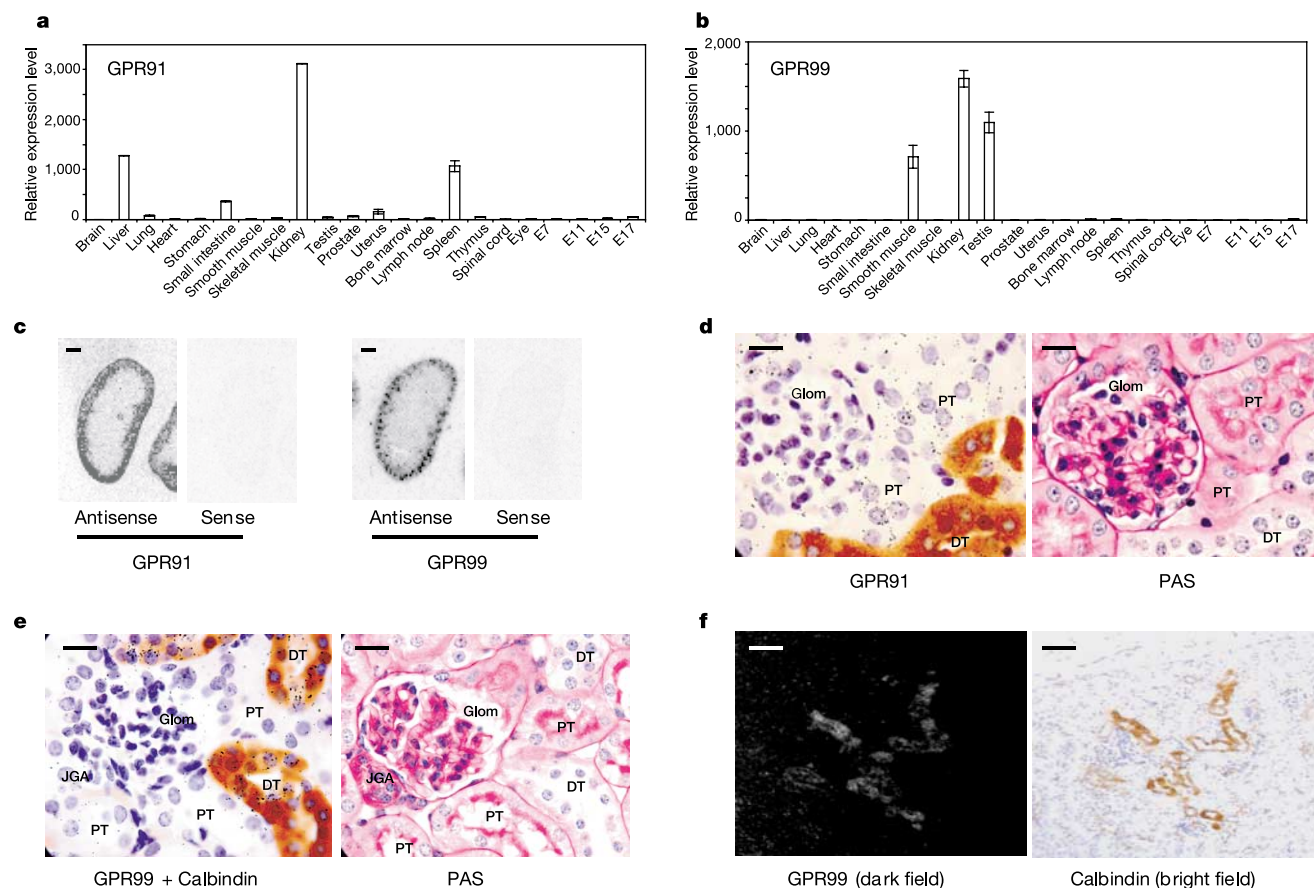


Figure 4 Regional distribution of GPR91 and GPR99. **a**, **b**, Q-RT-PCR analysis of GPR91 (**a**) and GPR99 (**b**) in mouse tissue. Data are shown as means $\pm$ s.e.m. for triplicate determinations. **c**, Localization of GPR91 and GPR99 in mouse kidney sections by *in situ* hybridization. Scale bar, 1 mm. **d**, Localization of GPR91 (black grains) in the proximal tubules. PT, proximal tubule; DT, distal tubule; glom, glomerulus. Sections were

prestained with anti-calbindin (brown) before hybridizing to GPR91 probe. Periodic acid-Schiff (PAS) staining identified proximal tubules. Scale bar, 25 μ m. **e**, **f**, Localization of GPR99 (black grains in **e** and white signals in **f**) in the distal tubules. Scale bar: 25 μ m (**e**); 0.25 mm (**f**).

and may provide a unique electrostatic complementary binding environment for succinate (Fig. 3d). A comparison of about 300 family A GPCRs revealed that GPR91 and GPR99 uniquely contain all four basic residues, suggesting that GPR91 and GPR99 are specialized receptors for dicarboxylate ligands.

Expression analysis by quantitative reverse-transcriptase-mediated polymerase chain reaction (Q-RT-PCR) revealed that the GPR91 and GPR99 messenger RNAs were each predominantly expressed in the kidney, with limited expression in other tissues (GPR91 in the liver and the spleen, and GPR99 in the testis and the smooth muscle; Fig. 4a, b and refs 3, 4). *In situ* hybridization experiments indicated specific GPR91 and GPR99 expression in the cortical region of the mouse kidney (Fig. 4c). GPR91 was detected mainly in the proximal tubules, as demonstrated by the presence of a prominent brush border projecting into the lumen on periodic-acid-Schiff-stained adjacent sections (Fig. 4d). GPR91 was also found, to a smaller extent, in the distal tubules (revealed by staining with antibodies against the distal-tubule marker calbindin¹⁰; Fig. 4d) and the juxtaglomerular apparatus (Supplementary Information). In contrast, GPR99 was found predominantly in the distal tubules (Fig. 4e, f).

Intravenous injection of succinate into Sprague-Dawley rats increased plasma renin activity (Fig. 5a), which is consistent with a previous report that succinate-treated kidney cultures *ex vivo* released renin¹¹. Renin is a key enzyme in the renin-angiotensin

system, which is essential in blood pressure regulation^{12,13}. Cumulative intravenous administration of succinate produced a dose-dependent increase in mean arterial pressure (MAP) (Fig. 5b). Bilateral nephrectomy completely abolished this effect (Fig. 5c), demonstrating the kidney as the primary site of action. To dissect the pharmacological mechanism for succinate-induced hypertension, the angiotensin-converting enzyme inhibitor captopril was used. Pretreatment with captopril significantly attenuated the hypertensive effect of succinate (Fig. 5d), indicating that the angiotensin-converting enzyme product(s) might mediate hypertension induced by succinate. In contrast, neither nephrectomy nor captopril treatment had any effect on angiotensin II-induced hypertension (Fig. 5c, d).

To further explain the mechanisms underlying succinate-induced hypertension, we generated GPR91-deficient mice by homologous recombination (Fig. 5e, f). Mice homozygous for the targeted allele of GPR91 are viable and have no discernable phenotype. The systolic blood pressures and heart rates were similar in unanaesthetized wild-type and GPR91-deficient mice (Fig. 5g; systolic blood pressures were 129.5 ± 1.5 mmHg for wild-type and 131.2 ± 2.4 mmHg for GPR91-deficient mice; heart rates were 613 ± 12.4 beats min^{-1} for wild-type and 606 ± 11.4 beats min^{-1} for GPR91-deficient mice). However, succinate could no longer induce hypertension in GPR91-deficient mice (Fig. 5h, i). In contrast, angiotensin II increased blood pressure similarly in wild-type and

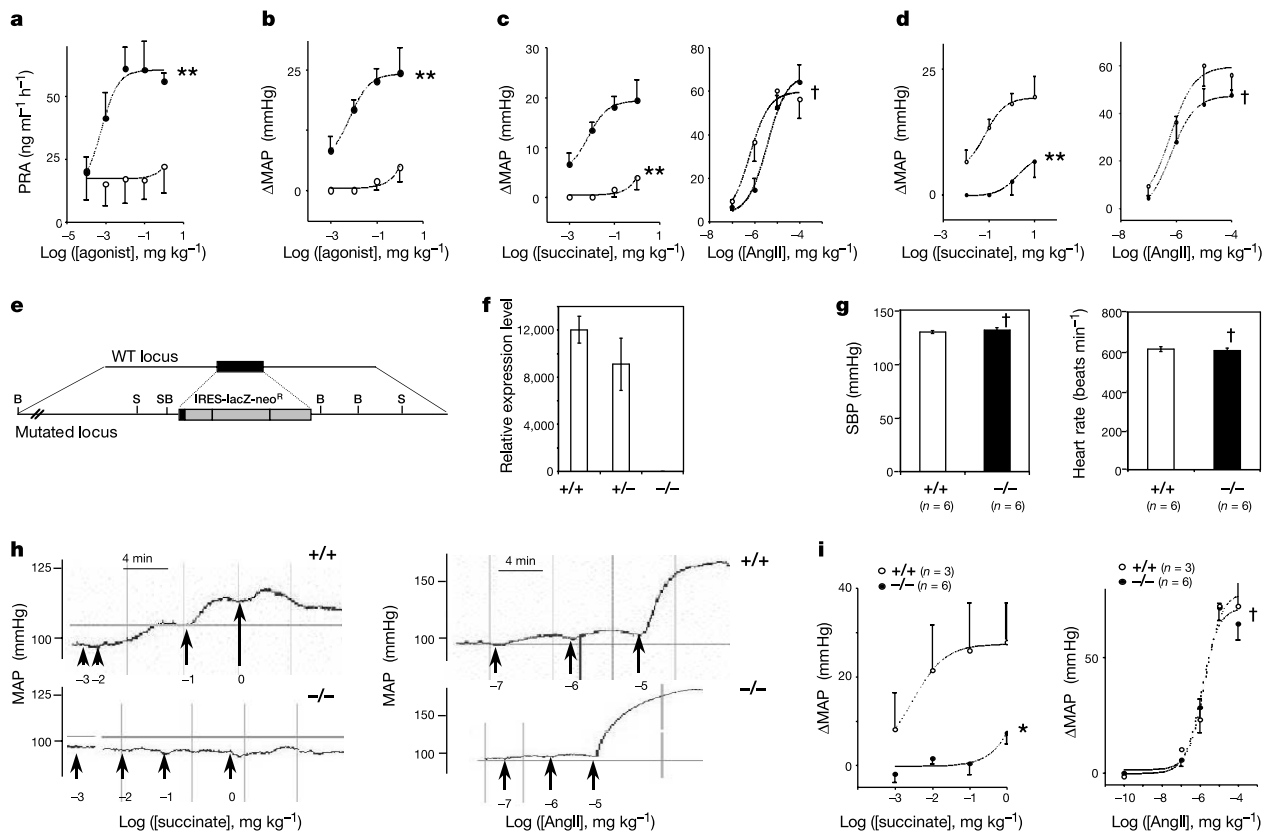


Figure 5 The pro-hypertensive effect of succinate and GPR91. All results are shown as mean $\pm$ s.e.m. Single asterisk, $P < 0.05$; double asterisk, $P < 0.01$; dagger, $P > 0.05$ (two-way analysis of variance or *t*-test). **a**, Plasma renin activity (PRA) in Sprague-Dawley rats by intravenous succinate infusion. Filled circles, succinate ($n = 6$); open circles, α -ketoglutarate ($n = 6$). **b**, MAP in Sprague-Dawley rats by intravenous succinate infusion. Values are the change in MAP (Δ MAP, mmHg) for each test group above its vehicle-injected control group. Filled circles, succinate ($n = 8$); open circles, α -ketoglutarate ($n = 6$). **c**, Effect of bilateral nephrectomy. Filled circles, kidney intact ($n = 6$); open circles, bilateral nephrectomy ($n = 4$). **d**, Effect of captopril. Filled circles,

captopril present ($n = 4$); open circles, captopril absent ($n = 6$). **e**, Schematic map of the GPR91 wild-type (WT) locus and the inactivated GPR91 allele. Black box, open reading frame of GPR91 gene; S, *SpeI*; B, *BamHI*; lacZ, β -galactosidase expression cassette; neo^R, neomycin resistance gene cassette. **f**, Q-RT-PCR analysis of GPR91 mRNA in kidneys of wild type (+/+), GPR91-heterozygous (+/-) and GPR91-deficient (-/-) mice. **g**, Systolic blood pressure and heart rate of unanaesthetized wild-type and GPR91-deficient mice. **h**, i, MAP tracing measurements (**h**) and dose-response curves (**i**) of wild-type and GPR91-deficient mice administered intravenously with succinate (left panels) or angiotensin II (AngII) (right panels).

GPR91-deficient mice (Fig. 5h, i), indicating that the hypertensive effect of succinate might be mediated by GPR91.

Succinate and α -ketoglutarate are normally present in mitochondria, but are found in circulation with mean plasma levels of about 5 μ M for succinate and about 25 μ M for α -ketoglutarate^{14–16}. Succinate is known to increase the reabsorption of phosphate and glucose into the proximal tubule¹⁷ and to stimulate gluconeogenesis¹⁸. Accumulation of extracellular succinate is observed in physiological and/or pathophysiological states linked to the deterioration of blood supply, such as ischaemia^{19,20}. It is known that the renovascular hypertension is often associated with restricted blood supply to the kidney owing to the obstruction of renal arteries in diseases such as atherosclerosis⁵. It is possible that the accumulation of succinate in ischaemic conditions might contribute to stenosis-associated hypertension through the activation of GPR91. Although we have shown that the hypertensive effect of succinate is mediated by GPR91 through the activation of the renin–angiotensin system, the molecular mechanism of renin release by succinate is not clear. Because GPR91 is a G_q - and G_i -coupled receptor, it is unlikely that GPR91 functions directly on the juxtaglomerular cells to stimulate renin release, which is usually triggered by G_s activation²¹. It remains to be investigated whether sympathetic nerve activity, prostaglandins or nitric oxide²² might provide an indirect mechanism for succinate-induced renin release. The presence of GPR91 in the liver and the spleen might also have functions in mediating these effects.

Thus, we have identified succinate as the natural ligand for GPR91, and α -ketoglutarate as the ligand for the homologous receptor GPR99. We have shown that intermediates in the citric acid cycle function as signalling molecules, engendering renewed interest in a biochemical pathway discovered more than 60 years ago. We have provided evidence that the pro-hypertensive effect of succinate is mediated by GPR91 and involves the renin–angiotensin system. GPR91 activation by succinate might therefore provide a link between energy homeostasis/metabolic status and haemodynamic regulation. The finding that intermediates in the citric acid cycle are GPCR ligands should facilitate the understanding of molecular links of the citric acid cycle to metabolic diseases, such as hypertension, atherosclerosis and diabetes, and the design of novel drugs with GPR91 and GPR99 as molecular targets. □

Methods

Purification of GPR91 ligand

Pig kidneys (6 kg; Pel-Freeze) were homogenized and extracted in ethanol/water/acetic acid (50/46/4, by vol.). The supernatant was freeze-dried, resuspended in 10 mM K_2HPO_4 pH 8.0, applied to an XK50/20 Q Sepharose anion-exchange column (Amersham), and developed with a 0–0.25 M NaCl linear gradient. The active fractions were subjected to three more steps of purification: a Hiload 26/60 Superdex30 size-exclusion column (developed in Hank's buffered saline at pH 7.35); a Superdexpeptide HR10/30 size-exclusion column (developed in 0.1% trifluoroacetic acid/water); and a C_{18} 4.6 mm $\times$ 250 mm reverse-phase column (developed in 0.1% trifluoroacetic acid/water). The active fractions from the final purification step were freeze-dried and dissolved in H_2O for NMR and mass spectrometry analysis.

NMR and mass spectrometry

The 1H and ^{13}C NMR spectra were recorded on a Bruker Avance DRX700 spectrometer (Bruker Analytik GmbH) at 700 and 176 MHz, respectively. The mass spectrum was obtained on an HP 1100MSD spectrometer (Agilent Technologies) with electrospray ionization in positive detection mode. A full-scan mass spectrum covering the m/z range 50–500 was acquired.

Cell culture and signalling assays

Chinese hamster ovary (CHO) cells were transiently transfected with 5 μ g GPCR and 5 μ g aequorin reporter gene² in 100-mm dishes as described²³. Aequorin luminescence was recorded with a Microumat (Berthold). FLIPR assays were performed on FLIPR384 with a FLIPR calcium assay kit (Molecular Devices). Inositol phosphate formation was assayed as the incorporation of 3H -myo-inositol as described^{23,24}. Cells were stimulated with compounds in the presence of 10 mM LiCl for 40 min at 37 °C before the addition of 20 mM formic acid. cAMP assay was performed with the cAMP-Screen system (Applied Biosystems). Cells were stimulated for 20 min with 10 μ M forskolin before treatment with compounds for a further 20 min at 37 °C. Anti-phospho-Erk and anti-Erk monoclonal antibodies were obtained from New England Biolabs. Anti-Flag and anti-tubulin

antibodies were obtained from Sigma. Western blots were detected with the ECL Plus detection system (Amersham). For some studies, 100 ng ml⁻¹ pertussis toxin (Calbiochem) was incubated with cells for 16 h.

Immunofluorescence staining

Human embryonic kidney 293 cells transfected with Flag-tagged GPR91 or GPR99 were plated on poly-(D-lysine)-treated slides. Cell-surface staining and ligand-induced internalization was performed as described⁹. Internalization of GPR91 or GPR99 was induced for 10 min by succinate (200 μ M) or α -ketoglutarate (300 μ M) at 37 °C.

In situ hybridization, immunohistochemistry and expression analysis

These were performed as described^{10,25}. [³³P]UTP-labelled antisense or sense RNA probes for mouse GPR91, GPR99 or renin were hybridized to paraformaldehyde-fixed, paraffin-embedded mouse kidney sections. For periodic acid–Schiff and renin staining, adjacent mouse kidney sections were used. For co-localization with calbindin on the same tissue section, slides were incubated with anti-calbindin antibody (Sigma) and developed with Vectastain avidin–biotin–complex kit (Vector Lab) with 3,3'-diaminobenzidine as substrate before hybridization *in situ*. Sections were counterstained with haematoxylin or toluidine blue. Q-RT-PCR was performed on an ABI Prism 7700 sequence detector using Taqman Gold kit (Applied Biosystems). Ratios of GPR91 or GPR99 to glyceraldehyde-3-phosphate dehydrogenase mRNA were calculated. Primer and probe sequences for GPR91 and GPR99 are available from the authors on request.

Receptor model

The published bovine rhodopsin structure was used as a template for generating a partial GPR91 structure model²⁶. The GPR91 primary sequence was aligned to bovine rhodopsin (PDB code 1F88; Supplementary Information), and the three-dimensional homology model was generated with the Modeller program.

In vivo experiments

All experiments were performed under protocols approved by the Institutional Animal Care and Use Committee. Animals were housed in a temperature-controlled room under a 12 h light–dark cycle with water and standard rodent chow available *ad libitum*. Blood pressure was measured in unanaesthetized animals by the tail-cuff method or in pentobarbital-anaesthetized animals by the intra-arterial method for those involving chemical administration^{27,28}. Male Sprague–Dawley rats (about 300 g; Harlan) or mice (about 30 g; Lexicon) were used. Systolic blood pressure was recorded in mice for 5 days; at least 20 measurements were made daily with a BP-2000 blood pressure analysis tail-cuff system (Visitech Systems). The mean systolic blood pressure and pulse rate were taken for each animal. Intra-arterial measurement of blood pressure was achieved with a polyethylene PE50 (for rats) or a stretched PE10 (for mice) catheter implanted into the carotid artery, with another implanted into the jugular vein before the experiments. The artery catheter was connected to a high-sensitivity isometric transducer (Harvard Apparatus) for continuous recording of blood pressure. MAP data were acquired by Ponemch Physiology Platform (Gould Instrument). The venous catheter was used for chemical administration. Bilateral nephrectomy was performed as described²⁹. Captopril (10 mg kg⁻¹) was given as a bolus injection 30 min before the administration of succinate. Plasma renin activity was determined by radioimmunoassay as the rate of angiotensin I generation (Diasorin). GPR91-deficient mice were generated by homologous recombination in embryonic stem cells (Lexicon). Results are expressed as means $\pm$ s.e.m. Statistical significance was evaluated by Student's *t*-test or two-way analysis of variance, followed by Dunnett's *post-hoc* test. Differences resulting in $P < 0.05$ were considered significant.

Received 3 February; accepted 15 March 2004; doi:10.1038/nature02488.

- Krebs, H. A. The history of the tricarboxylic acid cycle. *Perspect. Biol. Med.* **14**, 154–170 (1970).
- Eng, C., Kiuru, M., Fernandez, M. J. & Aaltonen, L. A. A role for mitochondrial enzymes in inherited neoplasia and beyond. *Nature Rev. Cancer* **3**, 193–202 (2003).
- Wittenberger, T., Schaller, H. C. & Hellebrand, S. An expressed sequence tag (EST) data mining strategy succeeding in the discovery of new G-protein coupled receptors. *J. Mol. Biol.* **307**, 799–813 (2001).
- Wittenberger, T. *et al.* GPR99, a new G protein-coupled receptor with homology to a new subgroup of nucleotide receptors. *BMC Genomics* **3**, 17–22 (2002).
- Olin, J. W. Atherosclerotic renal artery disease. *Cardiol. Clin.* **20**, 547–562 (2002).
- Jandeleit-Dahm, K. & Cooper, M. E. Hypertension and diabetes. *Curr. Opin. Nephrol. Hypertens.* **11**, 221–228 (2002).
- Stables, J. *et al.* A bioluminescent assay for agonist activity at potentially any G-protein-coupled receptor. *Anal. Biochem.* **252**, 115–126 (1997).
- Lee, D. K. *et al.* Discovery and mapping of ten novel G protein-coupled receptor genes. *Gene* **275**, 83–91 (2001).
- von Zastrow, M. & Kobilka, B. K. Antagonist-dependent and -independent steps in the mechanism of adrenergic receptor internalization. *J. Biol. Chem.* **269**, 18448–18452 (1994).
- Campean, V., Kricke, J., Ellison, D., Luft, F. C. & Bachmann, S. Localization of thiazide-sensitive Na⁺–Cl⁻ cotransport and associated gene products in mouse DCT. *Am. J. Physiol. Renal Physiol.* **281**, F1028–F1035 (2001).
- Baumbach, L., Leyssac, P. P. & Skinner, S. L. Studies on renin release from isolated superfused glomeruli: effects of temperature, urea, ouabain and ethacrynic acid. *J. Physiol. (Lond.)* **258**, 243–256 (1976).
- Lavoie, J. L. & Sigmund, C. D. Minireview: overview of the renin–angiotensin system—an endocrine and paracrine system. *Endocrinology* **144**, 2179–2183 (2003).
- Gurley, S. B., Le, T. H. & Coffman, T. M. Gene-targeting studies of the renin–angiotensin system: mechanisms of hypertension and cardiovascular disease. *Cold Spring Harb. Symp. Quant. Biol.* **67**, 451–457 (2002).

14. Bergmeyer, H. U. *Methods of Enzymatic Analysis* (Verlag Chemie, Weinheim/Bergstr., 1974).
15. Kushnir, M. M., Komaromy-Hiller, G., Shushan, B., Urry, F. M. & Roberts, W. L. Analysis of dicarboxylic acids by tandem mass spectrometry. High-throughput quantitative measurement of methylmalonic acid in serum, plasma, and urine. *Clin. Chem.* **47**, 1993–2002 (2001).
16. Martin, M., Ferrier, B. & Bayerl, G. Transport and utilization of alpha-ketoglutarate by the rat kidney *in vivo*. *Pflügers Arch.* **413**, 217–224 (1989).
17. Gullans, S. R., Kone, B. C., Avison, M. J. & Giebisch, G. Succinate alters respiration, membrane potential, and intracellular K^+ in proximal tubule. *Am. J. Physiol.* **255**, F1170–F1177 (1988).
18. Gullans, S. R., Brazy, P. C., Dennis, V. W. & Mandel, L. J. Interactions between gluconeogenesis and sodium transport in rabbit proximal tubule. *Am. J. Physiol.* **246**, F859–F869 (1984).
19. Krebs, H. A. Rate control of the tricarboxylic acid cycle. *Adv. Enzyme Regul.* **8**, 335–353 (1970).
20. Hems, D. A. & Brosnan, J. T. Effects of ischaemia on content of metabolites in rat liver and kidney *in vivo*. *Biochem. J.* **120**, 105–111 (1970).
21. Pan, L. *et al.* Critical roles of a cyclic AMP responsive element and an E-box in regulation of mouse renin gene expression. *J. Biol. Chem.* **276**, 45530–45538 (2001).
22. Hackenthal, E., Paul, M., Ganten, D. & Taugner, R. Morphology, physiology, and molecular biology of renin secretion. *Physiol. Rev.* **70**, 1067–1116 (1990).
23. An, S. *et al.* Identification and characterization of a melanin-concentrating hormone receptor. *Proc. Natl Acad. Sci. USA* **98**, 7576–7581 (2001).
24. Brandish, P. E., Hill, L. A., Zheng, W. & Scolnick, E. M. Scintillation proximity assay of inositol phosphates in cell extracts: high-throughput measurement of G-protein-coupled receptor activation. *Anal. Biochem.* **313**, 311–318 (2003).
25. Chuang, P. T., Kawcak, T. & McMahon, A. P. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes Dev.* **17**, 342–347 (2003).
26. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000).
27. Kregel, J. H., Hodgins, J. B., Hagaman, J. R. & Smithies, O. A noninvasive computerized tail-cuff system for measuring blood pressure in mice. *Hypertension* **25**, 1111–1115 (1995).
28. Sugiyama, F. *et al.* QTL associated with blood pressure, heart rate, and heart weight in CBA/Cal and BALB/c mice. *Physiol. Genomics* **10**, 5–12 (2002).
29. Silva, A. P. *et al.* Bilateral nephrectomy delays gastric emptying of a liquid meal in awake rats. *Ren. Fail.* **24**, 275–284 (2002).
30. Woronicz, J. D., Gao, X., Cao, Z., Rothe, M. & Goeddel, D. V. I κ B kinase- β : NF- κ B activation and complex formation with I κ B kinase- α and NIK. *Science* **278**, 866–869 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Reagan for critical comments during ligand identification; L. Yang for sharing reagents for *in situ* hybridization; G. Cutler, J. Knop, H. Baribault, J. Ma, S.-C. Miao, W. Inman, C. Ogden, S. Shuttleworth and M. Rich for providing support and discussions; and D. Goeddel, B. Lemmon and T. Hoey for critical reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.L. (ling@tularik.com).

Aquaporin-0 membrane junctions reveal the structure of a closed water pore

Tamir Gonen¹, Piotr Sliz², Joerg Kistler³, Yifan Cheng¹ & Thomas Walz¹

¹Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

²Howard Hughes Medical Institute and Children's Hospital Laboratory of Molecular Medicine, 320 Longwood Avenue, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

³School of Biological Sciences, University of Auckland, Auckland PO Box 92019, New Zealand

The lens-specific water pore aquaporin-0 (AQP0) is the only aquaporin known to form membrane junctions *in vivo*¹. We show here that AQP0 from the lens core, containing some carboxy-terminally cleaved AQP0^{2,3}, forms double-layered crystals that recapitulate *in vivo* junctions. We present the structure of the AQP0 membrane junction as determined by electron crystallography. The junction is formed by three localized interactions

between AQP0 molecules in adjoining membranes, mainly mediated by proline residues conserved in AQP0s from different species but not present in most other aquaporins. Whereas all previously determined aquaporin structures show the pore in an open conformation^{4–9}, the water pore is closed in AQP0 junctions. The water pathway in AQP0 also contains an additional pore constriction, not seen in other known aquaporin structures^{4–9}, which may be responsible for pore gating.

AQP0 is a member of the aquaporin family, members of which form pores that are either highly selective for water (aquaporins) or also permeable to other small neutral solutes such as glycerol (aquaglyceroporins) (reviewed in ref. 10). To date, the atomic structures of three aquaporins have been determined (AQP1^{4–6}, GlpF^{7,8} and AQPZ⁹). Sequence alignment shows AQP0 to be closely related to the pure water channel AQP1 (43.6% identity, 62.6% similarity). The presence of His 172, a residue conserved only in aquaporins but substituted in aquaglyceroporins, also suggests that AQP0 forms a pure water pore. AQP0 water permeability at neutral pH is approximately 40 times lower than that of AQP1¹¹, but AQP0 water conductance doubles under mildly acidic conditions¹². In the case of aquaporins in plant roots, a pH-dependent closure of the water pores has been reported¹³. Thus, evidence suggests that certain aquaporin pores are gated.

AQP0 water pores are considered essential for the lens micro-circulation system, proposed to supply deeper-lying fibre cells with nutrients and to clear waste products^{14,15}. Unlike all other aquaporins, AQP0 is also present in membrane junctions. It is particularly enriched in the 11–13 nm thin junctions between lens fibre cells, that feature square AQP0 arrays¹. Atomic force microscopy analysis of *in vitro* reconstituted AQP0 two-dimensional crystals demonstrated these crystals to be double-layered¹⁶.

Using AQP0 from the core of sheep lenses, where some of the AQP0 is proteolytically cleaved near the C terminus at various sites in an age-dependent manner^{2,3}, we reproduced the double-layered two-dimensional crystals^{16,17}. When core AQP0 was reconstituted at a lipid-to-protein ratio of 0.25 (w/w), large membrane sheets formed (>6 μ m) that in some cases showed two parallel edges, revealing them to be double-layered (Fig. 1a). The crystals showing p422 symmetry had lattice constants of $a = b = 65.5$ Å and a thickness of 11 nm (Fig. 2a), the same dimensions as thin junctions in the lens¹. Double-layered AQP0 two-dimensional crystals are therefore likely to recapitulate thin lens fibre cell junctions.

Electron diffraction analysis of AQP0 crystals (tilted to an angle of up to 70°) produced strong diffraction spots to 3 Å resolution in all directions (Fig. 1b, c; the electron crystallographic data are summarized in Table 1). As the crystal structure of the homologous bovine AQP1 was available⁵, we determined the structure of the AQP0 membrane junction by molecular replacement, thus avoiding the cumbersome and time-consuming process of collecting high-resolution images of tilted specimens. Sequencing of cloned sheep AQP0 showed an identical amino acid sequence to bovine AQP0¹⁸, with the exception of three conservative (S20T, M90V and S240T) and one non-conservative substitutions (C14F).

Our model (Fig. 2a) shows unique features that enable AQP0 to form membrane junction interactions. These differ from those previously suggested based on atomic force microscopy data¹⁶. The extracellular surface of AQP0 is rather flat and the interactions are mediated by direct contacts of the corresponding loops in the opposing AQP0 molecules (Fig. 2a). Loop C, connecting α -helices three and four, is significantly shorter than in AQP1 and GlpF. The shortened loop C (also seen in AQP2, AQP5, AQP6 and AQP8) is crucial for the formation of the very tight AQP0 junction, as it allows three specific interactions to be formed that are mediated almost exclusively by proline residues.

The most striking interaction involves Pro 38, in extracellular loop A. The proline residues (Pro 38) from eight symmetry-related AQP0 molecules in the stacked tetramers come together to form a

unique rosette-like structure in the very centre of the junction (Fig. 2b, c). The two other interactions involve three AQP0 molecules that we designated A, B and C (Fig. 2d, e). The contacts are formed by residues in the shortened loop C, namely a Proline–Proline motif (Pro 109 and Pro 110), which is part of a one-turn helix (helix HC), and residues Arg 113 and Pro 123. The proline–proline motif in molecule A, located in the top layer of the junction, interacts with the symmetry-related proline–proline motif of mole-

cule C in the bottom layer. Arg 113 in loop C of molecule C is in close proximity with the carbonyl of Pro 123 of B, back in the top layer, making it a likely hydrogen-bonding partner. The connections continue in this way, such that Arg 113 of B interacts with Pro 123 of C, and the proline–proline motif of B interacts with the symmetry-related proline–proline motif in the next AQP0 molecule. Thus, all subunits in interacting AQP0 tetramers are connected to two other subunits in the opposite layer.

In top views, the water pore can readily be seen in AQP1 (Fig. 3a), whereas the AQP0 pore appears much more constricted (Fig. 3b). About half of the residues lining the pores differ between AQP0 and AQP1, and many residues in AQP1 are substituted with larger and more hydrophobic ones in AQP0. The AQP1 pore profile, calculated with the program 'HOLE'¹⁹, shows a single constriction site formed by residues Phe 58, His 182, Arg 197 and Cys 191 (left and middle panel Fig. 3c). This ar/R site (so called because it contains an aromatic and an Arg residue)²⁰ was arbitrarily chosen as pore height 0. The ar/R site in AQP0 (left and right panel Fig. 3c) is formed by the corresponding residues Phe 48, His 172, Arg 187 and Ala 181, but it shows noteworthy differences. Ala 181 replaces AQP1's Cys 191, thus rendering AQP0 water conductance insensitive to mercurials²¹. His 172 moves towards the centre of the pore compared with AQP1's His 182, resulting in a minimum pore diameter of 1.96 Å, which is too narrow for water permeation. Moreover, the conformation of Arg 187 differs from those of the corresponding arginine residues in the ar/R constriction site in all other aquaporin structures. The distinctive conformation of Arg 187 is stabilized by hydrogen bonds with the backbone carbonyl of Ala 117 and with the side chain of Asn 119 of loop C. Whereas the ar/R constriction in AQP1 is only 1 Å long, constriction site I in AQP0 has a length of approximately 10 Å. The extension of the constriction site towards the extracellular surface (Fig. 3c, pore height 0 to 5 Å) results from the orientation of residues Ser 31 and Met 176 in the extracellular vestibule, whereas the extension of the constriction into the pore (Fig. 3c, pore height 0 to –5 Å) is due to residues Tyr 23, His 172, Asn 184 (part of the NPA motif in loop E) and Gly 182.

The AQP0 pore is also significantly narrower than the AQP1 pore at the site of the NPA motif in the cytoplasmic half of the pore (Fig. 3c, pore height –11 Å), resulting from the side chain of Phe 141 (Leu 151 in AQP1) pointing towards Asn 68 of the NPA motif in loop B.

Constriction site II in the cytoplasmic half of the pore (Fig. 3c, pore height –18 Å) is new. The side chain of Tyr 149 extends into the water pathway and together with Phe 75 and His 66 constricts the pore to a minimum diameter of 2.0 Å, again too narrow for water to pass. It thus seems that on junction formation, AQP0 ceases to function as a water pore. We note, however, that our resolution is

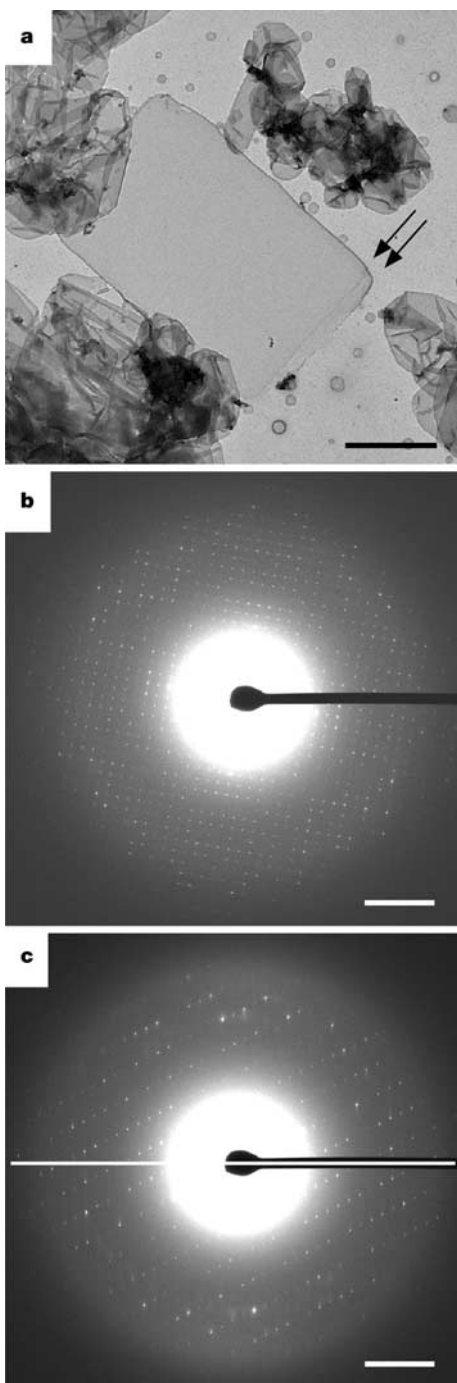


Figure 1 Double-layered two-dimensional crystals of AQP0. **a**, Reconstitution of lens core AQP0 produced large double-layered two-dimensional crystals. The edges of the two layers can be seen (arrows). **b**, Typical electron diffraction pattern of an untilted two-dimensional crystal showing diffraction spots to a resolution of 3 Å. **c**, Representative diffraction pattern of a two-dimensional crystal tilted to 70° showing strong and sharp diffraction spots, even perpendicular to the tilt axis. White line represents the orientation of the tilt axis. Scale bars correspond to 1 μm in **a** and (25 nm)^{–1} in **b** and **c**.

Table 1 Electron crystallographic data

Measurement	Value
Two-dimensional crystals	
Layer group	<i>p</i> 422
Unit cell	<i>a</i> = <i>b</i> = 65.5 Å
Thickness (assumed)	160 Å
Electron diffraction	
Number of patterns merged	131 (0°, 4°, 20°, 12°, 45°, 53°, 60°, 39°, 70°, 23°)
Resolution	3.0 Å in membrane plane, 3.5 Å perpendicular to membrane plane
<i>R</i> _{Friedel}	15.7%
<i>R</i> _{merge}	16.0% (54% at 3.5–3.0 Å)
Observed amplitudes to 3.0 Å	51,282
Maximum tilt	71.3°
Fourier space sampled	88.1% (82.0% at 3.5–3.0 Å)
Multiplicity	6.7 (4.5 at 3.5–3.0 Å)
Crystallographic refinement	
Crystallographic <i>R</i> -factor	30.20%
Free <i>R</i> -factor	33.85%
Ramachandran plot (%)	81.5/14/4/0.5 (favourable, additional, generous, disallowed)

currently limited to 3 Å, and even if a pore appears to be in a closed conformation, it might still be permeable to solutes²². Nonetheless, the length of constriction site I combined with the additional constriction site II makes it unlikely that the AQP0 pore in this conformation has significant water permeability.

Tyrosine 149, part of constriction site II, displays a relatively high *B*-factor, indicative of high mobility. Its hydroxyl group is also not hydrogen bonded and the side chain has sufficient space to change its conformation. It is therefore likely that Tyr 149 can swing out to open the water pore, thus constituting a possible gating mechanism

for the AQP0 pore. AQP0 water permeability is pH-dependent with a maximum at a slightly acidic pH¹². Two histidine residues, His 40 and His 66, are at positions where they could easily influence water conductance (Fig. 3c). Histidine 40, which has already been implicated in the pH-dependence of AQP0 water conductance¹², is situated right at the extracellular entrance of the pore, and the side chain extends into the water pathway. Histidine 66 is part of constriction site II. It is therefore likely that AQP0 water permeability may be affected by protonation/deprotonation of these two histidines.

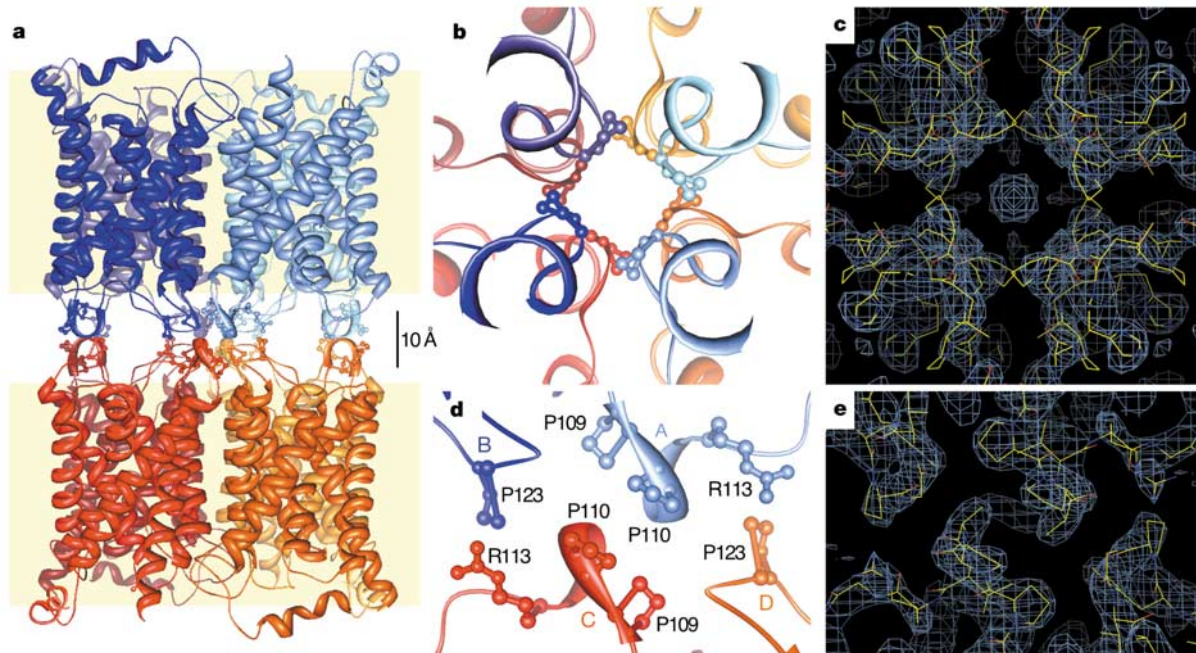


Figure 2 The AQP0-mediated membrane junction. **a**, Ribbon representation of the membrane junction with the positions of the two membranes indicated in yellow. **b**, The Pro 38 residues contributed by all eight AQP0 subunits in the two interacting tetramers. **c**, Corresponding area of the $2F_o - F_c$ map. **d**, The C loops connect each AQP0 molecule

to two molecules in the opposite membrane. Letters A to D refer to the molecules described in the text. **e**, Corresponding area of the $2F_o - F_c$ map. Panels **a**, **b** and **d** were created with Chimera²⁹ and panels **c** and **e** with the program O²⁸.

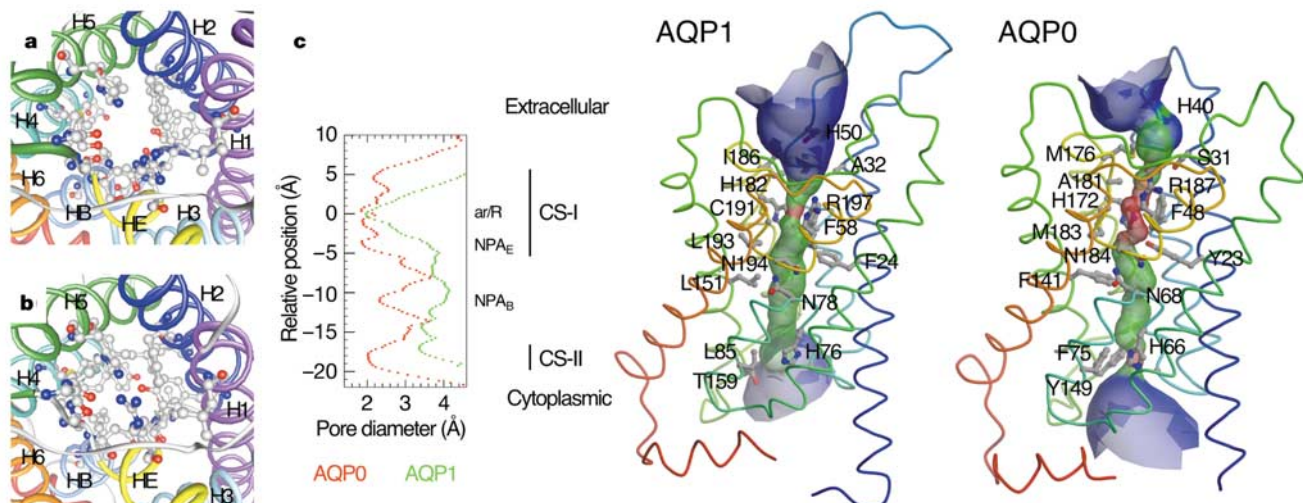


Figure 3 The closed AQP0 water pore. **a**, Side chains lining the pores in AQP1 (**a**) and AQP0 (**b**). **c**, The left panel shows pore profiles for AQP0 (red) and AQP1 (green) generated with the program 'HOLE'. The positions of the NPA motifs, the ar/R constriction site in

AQP1 and the two constriction sites in AQP0 (CS-I, CS-II) are marked and shown in red. The pore profiles are aligned with the models of AQP1 (middle panel) and AQP0 (right panel). Panels **a** and **b** were created with Chimera²⁹.

AQP0 forms functional water pores¹¹ with a maximal water conductance around pH 6.5¹². Our structure of junctional AQP0 determined from two-dimensional crystals produced at pH 6 shows the pore in a closed conformation. We propose that junction formation induces the closed pore conformation in our AQP0 structure. AQP0 and other aquaporins may be in a dynamic equilibrium between an open and a closed water pore conformation, a well-established concept in the ion channel field. Junction formation might then simply stabilize AQP0 in the closed pore conformation, which is otherwise not favoured at pH 6. In this respect, it is notable that the conformation of Arg 187, part of the ar/R constriction site, is stabilized by a hydrogen bond with Asn 119 in loop C, the loop that mediates most of the junction-forming interactions. It is therefore conceivable that junction formation involving residues Pro 109, Pro 110, Arg 113 and Pro 123 in loop C moves Asn 119 into a position in which it can stabilize the alternative conformation of Arg 187. This initial effect would then induce further conformational changes in pore-lining residues ultimately resulting in complete pore closure including Tyr 149 in constriction site II. The structure thus suggests a testable model for the way in which junction formation induces closure of the AQP0 water pore. □

Methods

Sequence of sheep lens AQP0

RNA was isolated from homogenized lens cortical tissue using Trizol (Invitrogen) according to manufacturer's instructions. First strand synthesis was performed using the Superscript first strand synthesis system for polymerase chain reaction with reverse transcription (RT-PCR) (Invitrogen). The synthesized complementary DNA was used for PCR amplification with Taq polymerase (Qiagen) and 1 µM of each sense (5'-CAGGGACCCAGGCGTTGT-3') and anti-sense (5'-ATGGCAGGAGCAAAGGAG-3') primers designed using the available sequence for bovine AQP0 (GenBank accession number NM173937). Products were resolved on a 0.8% agarose gel. The DNA was extracted using the quick gel extraction kit (Qiagen) and sequenced at the Harvard Biopolymer facility.

Purification of AQP0

Lenses were dissected to separate the soft cortical tissue from the hard core. Membranes were prepared from the lens core as described²³. Membranes were solubilized with 1% decyl maltoside (DM) in 10 mM Tris at pH 8, for 30 min at 37 °C, and insoluble material removed by centrifugation at 110,000 g. Proteins were bound to a MonoQ column (Amersham) equilibrated with 0.3% DM in 10 mM Tris at pH 8, and AQP0 eluted with 200 mM NaCl. Pooled fractions were run over a Superose 12 column (Amersham) equilibrated with 0.3% DM in 10 mM Tris at pH 8 and 150 mM NaCl.

Reconstitution of purified lens AQP0 into lipid bilayers

Purified AQP0 was mixed with DM-solubilized dimyristoyl phosphatidyl choline (DMPC) at varying lipid-to-protein ratios. The mixture was placed in a dialysis button and the detergent removed by dialysis against 10 mM MES at pH 6, 50 mM MgCl₂, 150 mM NaCl, 5 mM dithiothreitol and 0.02% NaN₃ at room temperature.

Electron microscopy and data processing

Negatively stained samples, prepared and imaged as described¹⁷, were used to assess the outcome of reconstitution experiments. For cryo-electron microscopy, two-dimensional crystallization samples were mixed with an equal amount of 20% glucose and the suspension applied to molybdenum grids (provided by Y. Fujiyoshi) covered with a thin layer of carbon film. Grids were blotted to remove excess material, loaded onto a high-tilt cryo-transfer holder (Gatan) and transferred into a Philips Tecnai F20 equipped with a field emission electron source and operated at 200 kV. The sample was cooled to the temperature of liquid nitrogen, and low-dose electron diffraction data were recorded with a Gatan 2Kx2K slow-scan CCD camera using a camera length of 3,000 mm and a selected area diffraction aperture of 70 µm. Electron diffraction patterns were analysed as described²⁴.

Molecular replacement and model building

The structure of AQP0 was determined by molecular replacement in MOLREP²⁵, using as a search model the AQP1 structure⁵ lacking amino acid residues 36–47 and 122–134. Crowther fast cross-rotation function calculations identified an orientation of a single subunit in the asymmetric unit as the top solution with an RF signal twice the value of the second peak. The initially determined unit cell axis $a = b$ of 68 Å was refined over ± 6 Å in 0.2 Å increments by repeating the search and optimizing the signal. The top solution was identified as 65 Å. The translation function with the monomer in the refined unit cell gave a top solution with an R -factor of 52.5% and a correlation coefficient of 42.7% (average values for the top 50 translation peaks were 57.8% and 25.3%, respectively). In the initial rounds of refinement we evaluated both electron and X-ray scattering factors. As expected for 3 Å resolution data, there was little difference in the resulting refinement statistics²⁶. In

all subsequent refinement rounds we used X-ray scattering factors.

At this stage the residues in AQP1 that differed from those in the AQP0 sequence were replaced by alanine (or glycine if the substitution was to glycine). The model was refined using CNS version 1.1 (ref. 27). In a first step the rigid body refinement was repeated, varying the unit cell dimension in 0.2 Å increments. The R -free minimum showed that the cell axis had to be adjusted to 65.5 Å, and this value was used in further refinements. Subsequent model building was performed with the program O (ref. 28) using $2F_o - F_c$ density modified by solvent flipping (see Supplementary Fig. 1), and simulated annealing composite omit-maps. The model was refined by simulated annealing with a number of refinement parameters (starting temperature, cooling rate, stereochemical weight term, energy constant of dihedral restraints for α -helical regions and range around restrained angles for helical regions) optimized by running several refinement cycles with a wide range of starting values and ultimately selecting the ones that gave the best R -free. Calculations were performed on the HMS structural biology Linux grid. The final refinement statistics are presented in Table 1.

Received 30 December 2003; accepted 19 March 2004; doi:10.1038/nature02503.

- Costello, M. J., McIntosh, T. J. & Robertson, J. D. Distribution of gap junctions and square array junctions in the mammalian lens. *Invest. Ophthalmol. Vis. Sci.* **30**, 975–989 (1989).
- Takemoto, L., Takehana, M. & Horwitz, J. Covalent changes in MIP26K during aging of the human lens membrane. *Invest. Ophthalmol. Vis. Sci.* **27**, 443–446 (1986).
- Roy, D., Spector, A. & Farnsworth, P. N. Human lens membrane: comparison of major intrinsic polypeptides from young and old lenses isolated by a new methodology. *Exp. Eye Res.* **28**, 353–358 (1979).
- Murata, K. *et al.* Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605 (2000).
- Sui, H., Han, B. G., Lee, J. K., Walian, P. & Jap, B. K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878 (2001).
- Ren, G., Reddy, V. S., Cheng, A., Melnyk, P. & Mitra, A. K. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl Acad. Sci. USA* **98**, 1398–1403 (2001).
- Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
- Tajkhorshid, E. *et al.* Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**, 525–530 (2002).
- Savage, D. F., Egea, P. F., Robles-Colmenares, Y., O'Connell, J. D. III & Stroud, R. M. Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS Biol.* **1**, 334–340 (2003).
- Agre, P. *et al.* Aquaporin water channels—from atomic structure to clinical medicine. *J. Physiol. (Lond.)* **542**, 3–16 (2002).
- Chandy, G., Zampighi, G. A., Kreman, M. & Hall, J. E. Comparison of the water transporting properties of MIP and AQP1. *J. Membr. Biol.* **159**, 29–39 (1997).
- Nemeth-Cahalan, K. L. & Hall, J. E. pH and calcium regulate the water permeability of aquaporin 0. *J. Biol. Chem.* **275**, 6777–6782 (2000).
- Tournaire-Roux, C. *et al.* Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. *Nature* **425**, 393–397 (2003).
- Mathias, R. T., Rae, J. L. & Baldo, G. J. Physiological properties of the normal lens. *Physiol. Rev.* **77**, 21–50 (1997).
- Donaldson, P., Kistler, J. & Mathias, R. T. Molecular solutions to mammalian lens transparency. *News Physiol. Sci.* **16**, 118–123 (2001).
- Fotiadis, D. *et al.* Surface tongue and groove contours on lens MIP facilitate cell-to-cell adherence. *J. Mol. Biol.* **300**, 779–789 (2000).
- Hasler, L. *et al.* Purified lens major intrinsic protein (MIP) forms highly ordered tetragonal two-dimensional arrays by reconstitution. *J. Mol. Biol.* **279**, 855–864 (1998).
- Gorin, M. B., Yancey, S. B., Cline, J., Revel, J. P. & Horwitz, J. The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning. *Cell* **39**, 49–59 (1984).
- Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
- de Groot, B. L. & Grubmüller, H. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**, 2353–2357 (2001).
- Varadaraj, K. *et al.* The role of MIP in lens fiber cell membrane transport. *J. Membr. Biol.* **170**, 191–203 (1999).
- Bond, P. J., Faraldo-Gomez, J. D. & Sansom, M. S. OmpA: a pore or not a pore? Simulation and modeling studies. *Biophys. J.* **83**, 763–775 (2002).
- Gonen, T., Donaldson, P. & Kistler, J. Galectin-3 is associated with the plasma membrane of lens fiber cells. *Invest. Ophthalmol. Vis. Sci.* **41**, 199–203 (2000).
- Mitsuoka, K. *et al.* The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution. *J. Mol. Biol.* **286**, 861–882 (1999).
- Vagin, A. & Teplyaev, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).
- Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421 (1996).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Huang, C. C., Couch, G. S., Pettersen, E. F. & Ferrin, T. E. Chimera: an extensible molecular modeling application constructed using standard components. *Pacif. Symp. Biocomput.* **1**, 724 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. C. Harrison for help with the molecular replacement, model building and writing of the manuscript. We also thank Y. Fujiyoshi, K. Mitsuoka and K. Tani for advice. This work was supported by National Institute of Health funding to T.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.W. (twalz@hms.harvard.edu). The sequence for sheep AQP0 has been deposited in GenBank (accession number AY573927). Coordinates and structure factors have been deposited in the Protein Data Bank (accession code 1SOR).

The GTPase-activating protein Rap1GAP uses a catalytic asparagine

Oliver Daumke*, Michael Weyand*, Partha P. Chakrabarti, Ingrid R. Vetter & Alfred Wittinghofer

Max-Planck-Institut für Molekulare Physiologie, Otto-Hahnstr. 11, 44227 Dortmund, Germany

* These authors contributed equally to this work

Rap1 is a Ras-like guanine-nucleotide-binding protein (GNBP) that is involved in a variety of signal-transduction processes^{1,2}. It regulates integrin-mediated cell adhesion and might activate extracellular signal-regulated kinase. Like other Ras-like GNBP, Rap1 is regulated by guanine-nucleotide-exchange factors (GEFs) and GTPase-activating proteins (GAPs). These GAPs increase the slow intrinsic GTPase reaction of Ras-like GNBP by many orders of magnitude and allow tight regulation of signalling. The activation mechanism involves stabilization of the catalytic glutamine of the GNBP and, in most cases, the insertion of a catalytic arginine of GAP into the active site³. Rap1 is a close homologue of Ras but does not possess the catalytic glutamine essential for GTP hydrolysis in all other Ras-like and Gα proteins. Furthermore, RapGAPs are not related to other GAPs and apparently do not use a catalytic arginine residue⁴. Here we present the crystal structure of the catalytic domain of the Rap1-specific Rap1GAP at 2.9 Å. By mutational analysis, fluorescence titration and stopped-flow kinetic assay, we demonstrate that Rap1GAP provides a catalytic asparagine to stimulate GTP hydrolysis. Implications for the disease tuberous sclerosis are discussed.

Rap1GAP (663 amino acids, molecular mass of 73 kDa) is the founding member of a family of GAPs specific for Rap1⁵, and a sequence of 340 amino acids within it is sufficient for Rap1GAP activity^{4,6}. Homologous sequences have been identified in the human genes *SPA1* (ref. 7), *E6TP1* (ref. 8) and *TSC2*, which encodes tuberin⁹. RapGAPs are thought to be important for tumour suppression because degradation of the E6TP1 protein via the transforming papilloma virus protein E6 correlates with development of cervical cancer¹⁰. Deletion of the *Spa1* gene in mouse creates a spectrum of myeloid disorders that resemble chronic myeloid leukaemia¹¹. Furthermore, mutations in the GAP domain of the tumour suppressor tuberin, a GAP for the Ras-like protein RheB, can cause the benign tumour phenotype tuberous sclerosis⁹.

We solved the structure of the catalytic domain of Rap1GAP (residues 75–415)^{4,6} by X-ray crystallography to a resolution of 2.9 Å, with an R_{work} of 23.2% and an R_{free} of 27.6% (see Methods, Supplementary Tables 1 and 2). Out of four molecules constituting two dimers (molecule A–B and molecule C–D) in the asymmetric unit, three are well defined in the electron density (molecules A, B

and C). Because no major difference was observed between molecules A and C (root-mean-square deviation 0.5 Å for the Cα atoms), the dimer A–B will be described here. Dimerization was also observed in size-exclusion chromatography in which Rap1GAP elutes with an apparent molecular mass of ~100 kDa. The size of the dimer interface is approximately 2,400 Å² and is formed mainly by hydrophobic residues.

Each monomer consists of two domains, which we call the dimerization and the catalytic domain. Both domains have an α–β fold, with a central β-sheet surrounded by α-helices (Fig. 1a, Supplementary Fig. 1). Remarkably, the carboxy-terminal catalytic domain shows structural similarity to the G domain of Ras superfamily proteins (Fig. 1b, Supplementary Fig. 2). Its long C-terminal helix, α9, folds back onto the amino-terminal dimerization domain. Because no extensive interaction is observed between the two domains we assume that their relative orientation is variable in solution. Residues outside the catalytic domain are much less conserved, and tuberin does not show any homology to the dimerization domain (Supplementary Fig. 1). Rap1GAP has no structural similarity to GAPs of Ras, Rho, Arf, Rab or Ran³.

To investigate the role of dimerization, we replaced two hydrophobic residues from the mostly hydrophobic dimerization interface with large polar residues. The resulting mutant (F100E, L173E), which is less stable than the wild-type protein, eluted primarily as a monomer in size-exclusion chromatography (data not shown). In a standard multiple turnover assay (excess Rap1·GTP compared with Rap1GAP, see Methods), this mutant had nearly wild-type activity (Fig. 2). This indicates that dimerization is not required for Rap1GAP activity.

As the sequence of the catalytic domain is more conserved than that of the dimerization domain (Supplementary Fig. 1) and contains two lysines essential for Rap1GAP activity⁴, we asked whether the catalytic domain alone is sufficient for Rap1GAP activity. However, constructs containing the core or a more extended catalytic domain were either insoluble or inactive (data not shown). This implies that both domains are necessary for GAP function. Because helix α9 interacts with the dimerization domain, we asked whether helix α9 supports Rap1GAP activity. Figure 2 shows that mutation of the invariant, surface-exposed Arg 388 in helix α9 (Fig. 1c) decreases Rap1GAP activity (see also ref. 4).

Mutations of the highly invariant lysine residues Lys 194 and Lys 285 reduce Rap1GAP activity by 25-fold and 100-fold, respectively, by primarily affecting the affinity for Rap1·GTP⁴. Contrary to expectation, neither of the lysines is exposed to solvent but instead involved in polar interactions that apparently stabilize the structural integrity around helix α7 (Fig. 1c). Lys 194 interacts with Asp 291 and Asn 290, which are residues of helix α7 that are conserved in the Rap1GAP family. Lys 285 of helix α7 makes contacts with the invariant Glu 207 and Ala 310. If the role of Lys 194 and Lys 285 is to position helix α7, then mutation of their contact residues should have a similar effect on catalysis (or binding). This is indeed confirmed by alanine mutations of Glu 207, Asp 291 and His 267, which reduce Rap1GAP activity depending on the loss of their hydrogen-bonding interactions (Fig. 2). The network of interactions described for Lys 194 and Lys 285 is only found in Rap1GAP monomers A and C, indicating a certain flexibility of helix α7 that might be relevant for the catalytic process. All the mutants described above show a drastically reduced affinity for Rap1 that is too low to be measured by either direct binding or multiple turnover K_m measurements. In common with the lysine mutants⁴, the major (and maybe exclusive) effect of these mutations is a reduction in affinity between Rap1·GTP and Rap1GAP.

We reasoned that the ionic interactions described above stabilize and position the putative interaction helix α7, which in turn might contain one or more residues essential for catalysis. The solvent-

exposed invariant residues in helix $\alpha 7$ (Fig. 1c) were thus mutated, and the activities of the R286A, H287A and N290A mutants were measured in the standard hydrolysis assay (Fig. 2). R286A showed a fourfold reduction in activity, which is similar to previous descriptions⁴. Replacement of the more buried His 287 reduces activity drastically. Finally, mutation of the surface-exposed Asn 290

completely eliminates GAP activity on Rap1.

To further understand the mechanistic consequences of the mutations, we used a previously established stopped-flow fluorescence assay to monitor single turnover GAP reactions¹². Using Rap1•GTP labelled with a fluorescence reporter and wild-type Rap1GAP, we observed an increase in fluorescence owing to com-

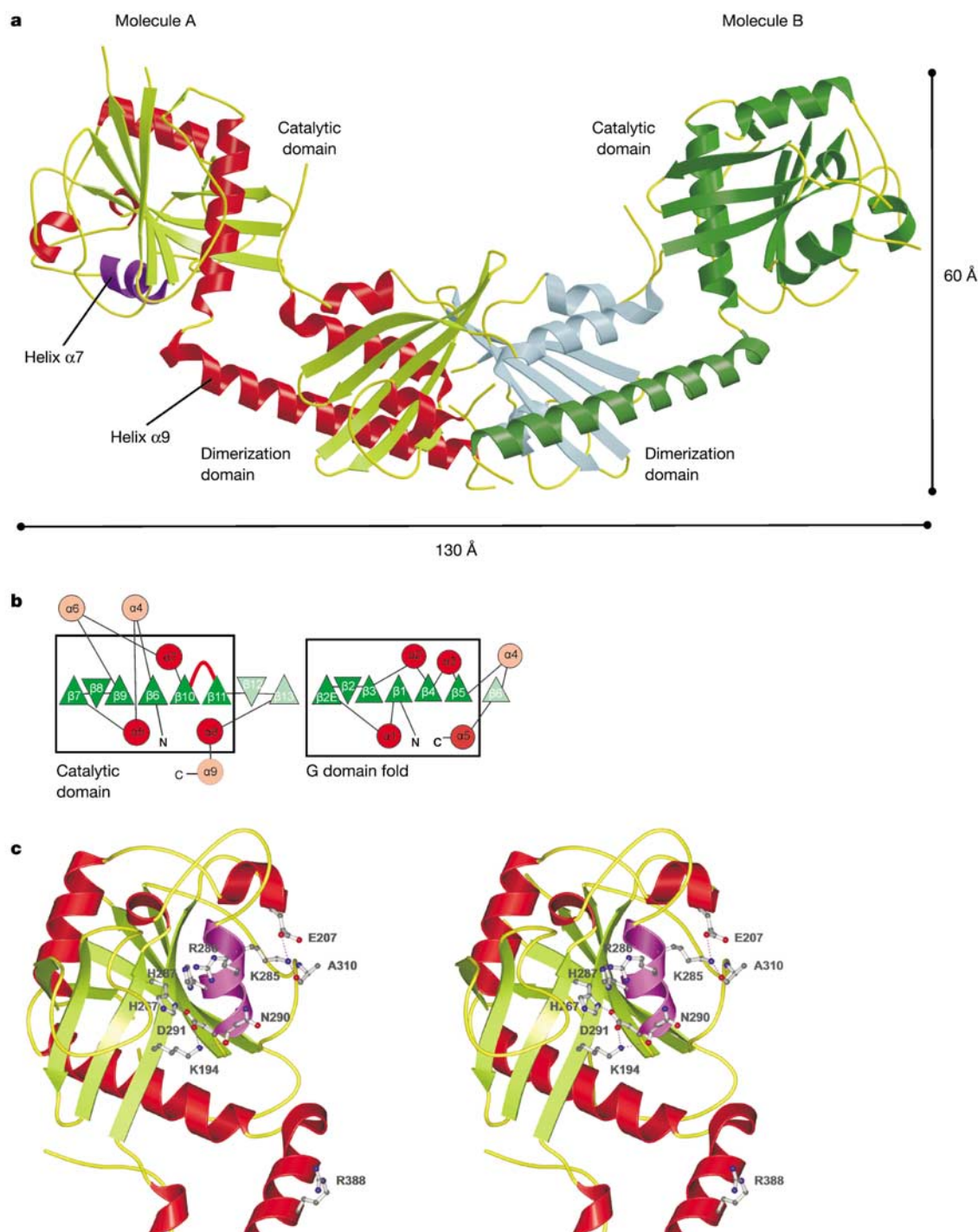


Figure 1 The structure of Rap1GAP. **a**, Ribbon-type presentation of the Rap1GAP dimer, in which molecule A is coloured according to the secondary structure and molecule B according to the domain definition (dimerization domain in light blue; catalytic domain in green). The Rap1 interaction helix $\alpha 7$ in molecule A is purple. **b**, Topology comparison of the Rap1GAP catalytic domain with those G domains that have an extra β -strand ($\beta 2_E$) in

the effector domain, such as Ran or Arf. Secondary-structure elements that can be superimposed in Rap1GAP and Ran are boxed (see also Supplementary Fig. 2). **c**, Stereo plot of the catalytic centre of Rap1GAP. Residues that are considered to be important for binding and/or catalysis are shown in ball-and-stick presentation. Dashed lines indicate hydrogen bonds or polar interactions.

plex formation followed by a decrease due to GTP hydrolysis and concomitant complex dissociation (Fig. 3a). The rate-limiting step under this condition has a k_{cat} of 2 s^{-1} and is believed to represent the release of inorganic phosphate¹². Similar to results from the multiple turnover experiment, the R286A mutant had a slightly decreased activity, indicating that although this arginine is exposed and close to the reaction centre, it does not play a significant role in catalysis, in contrast to the arginine fingers of Ras-, Rho- and RabGAPs³. No association with Rap1 was seen for the H287A mutant, which demonstrates that His 287 is mostly (if not exclusively) involved in binding or stabilization. However, for the N290A mutant, we observed a fluorescence increase due to complex formation, but no decrease during the lifetime of the experiment. Because there is no hydrolysis, the amplitude of the fluorescent transients can be used to determine the affinity between Rap1 and Rap1GAP (N290A), which turns out to be $3 \mu\text{M}$ (Fig. 3b). The affinity between Rap1-GTP and wild-type Rap1GAP has been determined to be $14 \mu\text{M}$ (ref. 12). Thus the affinity of the mutant is even higher than that of the wild type. This clearly implies that Asn 290 is directly involved in catalysis. Even prolonged incubation shows no GTPase activity exceeding the intrinsic reaction rate, indicating that the N290A mutation reduces GAP activity by 10^5 -fold.

Aluminium fluoride (AlF_3) along with GDP (or ADP) is a mimic of the transition state for many phosphoryl transfer enzymes such as myosin and G α subunits of heterotrimeric G proteins¹³. For Ras-like proteins, aluminium fluoride binds to Ras proteins only in the presence of their respective GAPs¹⁴, demonstrating that the active site of Ras-like proteins needs to be complemented by residues from GAP. As shown in Fig. 3c by monitoring the fluorescence increase, Rap1-GDP forms a complex with Rap1GAP, but only in the presence of aluminium fluoride. As observed for the interaction of Ras and Rho with their respective GAPs, the association with the GAP is slow. The R286A mutant also associates with GAP but the amplitude of the fluorescence change is smaller. However, no association is seen for the N290A mutant. This is consistent with the assumption that Asn 290 directly stabilizes the transition state of GTP hydrolysis.

Single point mutations in the RapGAP homology region of tuberin have been described previously^{15–17}. When mapped onto the structure of Rap1GAP, many mutations appear to cluster in or close to the hydrophobic core and probably destabilize the protein

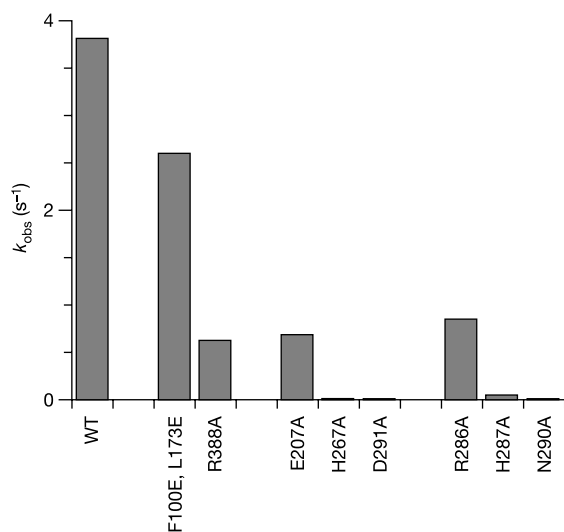


Figure 2 Mapping the active site. Bar chart of wild-type and mutant Rap1GAP activities, measured under standard multiple turnover condition. In case of drastically reduced activity the concentration of Rap1GAP was increased (see Methods).

(Fig. 4a). On the basis of the conservation of relevant residues, the mutations N1643K (refs 15, 16), N1643I (ref. 17) or R1743P (ref. 16) can most easily be explained by our findings. As Asn 1643 in tuberin is homologous to the catalytic asparagine in Rap1GAP, we

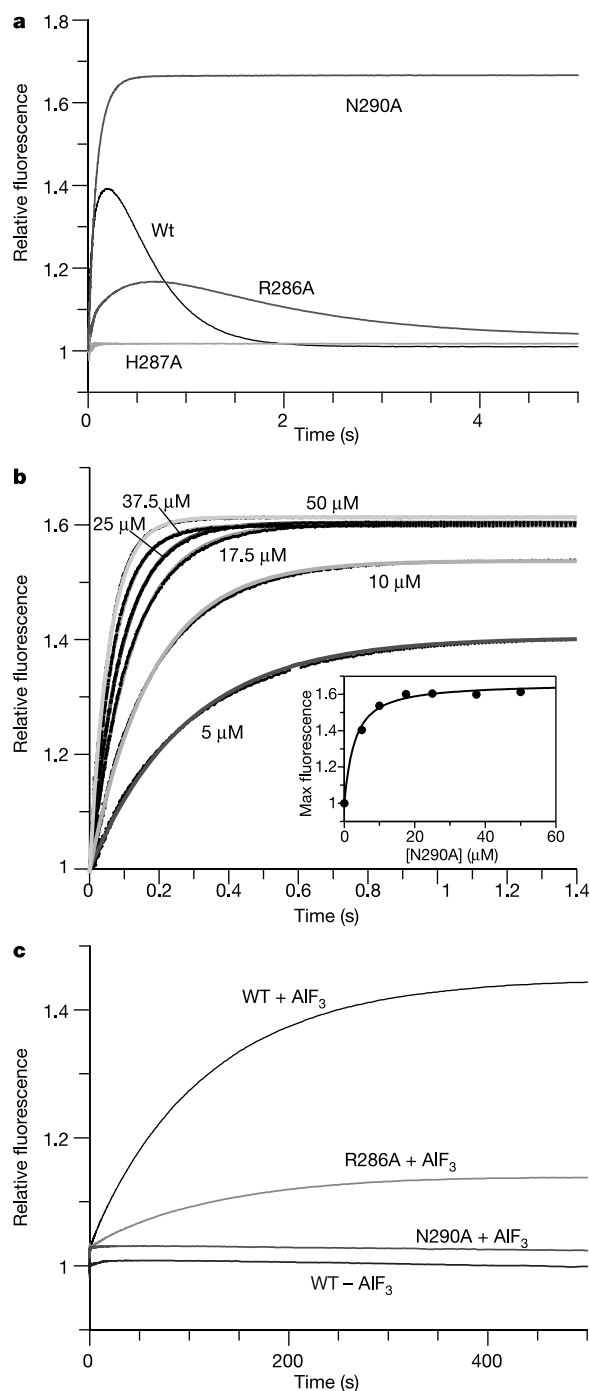


Figure 3 Mechanistic analysis of Rap1GAP catalysis. **a**, Stopped-flow fluorescence analysis of the reaction of Aedans-labelled Rap1-GTP with wild-type and mutant Rap1GAP (see Methods and ref. 12). The fluorescence transient of the wild type shows an increase due to complex formation and a decrease due to hydrolysis, whereas the N290A mutation shows only binding and no hydrolysis. **b**, The fluorescence amplitudes obtained with Aedans-labelled Rap1-GTP and increasing concentrations of Rap1GAP (N290A) are measured and plotted as a function of the latter concentration in the insert. Nonlinear data fitting results in a dissociation equilibrium constant of $3 \mu\text{M}$. **c**, Complex formation of Rap1-GDP and Rap1GAP in the presence or absence of AlF_3 is measured as an increase of fluorescence with time. Although the R286A mutant shows complex formation with decreased amplitude, no complex formation is observed for the N290A mutant.

made the corresponding mutations N290K and N290I in Rap1GAP and investigated the properties of the mutants. As expected, N290K is completely inactive (Fig. 4). Although the N290I mutation could not be investigated owing to solubility problems, it is also expected to be inactive. This indicates that, as in neurofibromatosis¹⁸, the mutation of the essential catalytic residue can induce the formation of benign tumours. In the case of tuberous sclerosis, it implies that upregulated RheB in the mTOR pathway contributes to unregulated growth⁹.

Arg 1576 of tuberin corresponds to the surface-exposed, highly conserved Arg 388 in Rap1GAP, which is located in helix $\alpha 9$ of the catalytic domain. Whereas the R388A mutant shows a fourfold reduction in affinity for Rap1 (Fig. 2 and ref. 4), the R388P mutant has an even more reduced activity (Fig. 4b). Because R388 is not in the direct vicinity of the phosphoryl transfer site, it is reasonable to assume that this residue contributes to the binding interface with Rap1 or RheB. Because a construct encompassing the catalytic domain of tuberin has no GAP activity on RheB or Rap1 (data not shown) under standard conditions, we assume that hamartin, which binds to tuberin and is required for its GAP activity⁹, plays a similar role to the dimerization domain in stabilizing the position of the helix $\alpha 9$ and forming an interface with the GTP-binding protein.

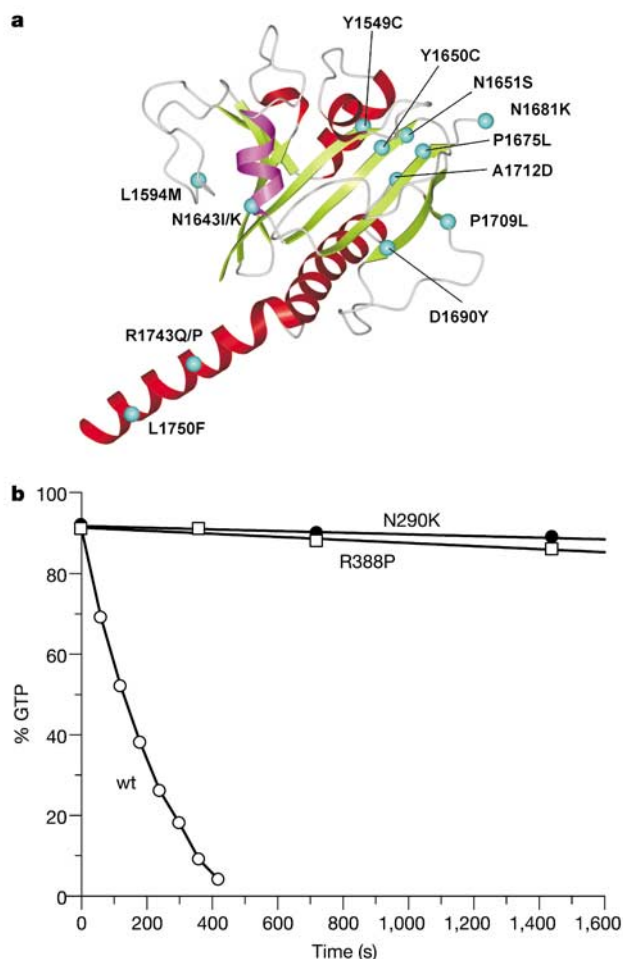


Figure 4 Implications for tuberin. **a**, Ribbon-type presentation of tuberin mutation sites, which are found in tuberous-sclerosis patients, based on the Rap1GAP structure (see Supplementary Fig. 1). **b**, Kinetic analysis of mutations that are analogous to those found in tuberous-sclerosis patients under standard conditions. Seventy nanomolar of Rap1GAP wild-type and of the R388P mutant were used and 60 μ M of the Rap1GAP (N290K) mutant.

GAP-stimulated GTPase reactions on most Ras-like proteins involve a glutamine residue, located on the GTP-binding protein, and an arginine residue that is supplied by the GAP. These two residues are of varying importance in the different systems. Mutation of the catalytic glutamine in Ras, Rho and Ran almost completely abolishes the GTPase reaction^{19,20}, whereas it can be partially tolerated in Rab²¹. The importance of the arginine finger is highest for RasGAP²² and RabGAP²³, somewhat lower for RhoGAP^{24,25}, and RanGAP does not employ an arginine finger²⁰. Its importance for ArfGAP is still debated²⁶. Here we have shown that in the absence of both a *cis*-glutamine and a *trans*-arginine, an asparagine, which we call the 'asparagine thumb' owing to its prominent location on the surface of helix $\alpha 7$, is absolutely essential for catalysis. As Gln 61 in Ras, Gln 63 in Rho and Gln 69 in Ran are involved in stabilizing the relative position of the nucleophilic water and the γ -phosphate^{24,27,28}, we assume that the carboxamide side chain of asparagine has a similar function. It is also remarkable that the catalytic domain of Rap1GAP displays high structural similarity to the Ras superfamily and Asn 290 is located in a helix topologically equivalent to the switch II helix $\alpha 2$ in the G domain (Fig. 1b, Supplementary Fig. 2), suggesting a common evolutionary origin. □

Methods

Structure determination

Purification, crystallization and data collection of native crystals are described elsewhere²⁹. Selenomethionine-substituted protein was purified and crystallized, and the crystals flash frozen in the same way as the non-substituted protein. Crystals belonged to the orthorhombic crystal system and contained four molecules in the asymmetric unit. Sixteen out of twenty selenium atoms were found by SHELXD (<http://shelx.uni-ac.gwdg.de/SHELXD/>) using the anomalous signal and the isomorphous signal of the selenomethionine-substituted crystals compared with wild-type crystals. Selenium sites were refined and initial phases were calculated using the program SHARP (<http://www.globalphasing.com/>). In the resulting electron density, the main chain of one molecule was traceable so that an initial Rap1GAP monomer model could be built. By using the non-crystallographic symmetry, derived from the selenium sites, we extended this initial model to the complete asymmetric unit. The model was built using the XtalView package (<http://www.scripps.edu/pub/dem-web/index.html>) and refined using Refmac5 with Translation, libration, screw-rotation displacement (TLS) refinement (<http://www.ysbl.york.ac.uk/~garib/refmac/index.html>). The final *R* values are *R*_{work} = 23.2% and *R*_{free} = 27.6%, respectively (see Supplementary Tables 1 and 2 for details). Two of the four molecules (molecules A and C) are defined clearly within the electron density. Molecule B is less well defined, and molecule D can be seen only partially. This enhanced flexibility of chain D is explained by the crystal packing because molecule D, apart from the dimerization interface, shows only contacts to other symmetry-related D molecules. This flexibility agrees with the moderate diffraction quality of the crystals. The final model contains 1,005 amino acids, 50 water, two sulphate and two 2-methyl-2,4-pentanediol molecules and has a good geometry with no residues in the disallowed region of the Ramachandran plot as judged by the program Procheck (<http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>) (Supplementary Fig. 3). Ribbon plots were prepared using Molscript (<http://www.avatar.se/molscript/>) and rendered with Raster3D (<http://www.bmsc.washington.edu/raster3d/raster3d.html>) and Povray (<http://www.povray.org>).

Biochemistry

Rap1B and Rap1GAPs were expressed and purified as described previously⁴. Mutants were prepared using the QuickChange protocol (Stratagene). Multiple turnover GTPase assays were performed at 20 °C with 100 μ M Rap1-GTP as the substrate and 70 nM Rap1GAP (or the indicated point mutants) as the enzyme and 50 mM HEPES (pH 7.5), 100 mM NaCl, 5 mM MgCl₂ as the reaction buffer. If no stimulation of GTP hydrolysis in Rap1 was observed, the GAP concentration was adjusted appropriately (300 nM for the E207A mutant, 8 μ M for the H267A mutant, 30 μ M for the D291A mutant, 4 μ M for the H287A mutant and 30 μ M for the N290A mutant). Initial rates of GAP-stimulated hydrolysis in Rap1-GTP were determined using standard HPLC measurement³⁰. Single turnover fluorescence assays were performed as described previously¹² at 10 °C using 50 mM HEPES (pH 7.5), 100 mM NaCl, 5 mM MgCl₂ as reaction buffer. Two micromolar Rap1-Aedans-GTP was 1:1 mixed with 50 μ M Rap1GAP or the indicated mutants in the stopped-flow apparatus, and the fluorescence was monitored. The fluorescence traces presented are the average of at least three individual measurements. Data processing was done as described previously¹² using the program Grafit 5 (Erithacus Software Limited). For AIF_x binding, Rap1-Aedans was incubated with a catalytic amount of Rap1GAP at 20 °C for 2 h to convert it to the GDP-bound form. Two micromolar Rap1-Aedans-GDP in 50 mM HEPES (pH 7.5), 100 mM NaCl, 5 mM MgCl₂, 500 μ M AlCl₃, 5 mM NaF were 1:1 mixed with 50 μ M of the corresponding Rap1GAP construct in the stopped-flow apparatus and the fluorescence monitored.

Received 12 February; accepted 16 March 2004; doi:10.1038/nature02505.

1. Bos, J. L., de Rooij, J. & Reedquist, K. A. Rap1 signalling: adhering to new models. *Nature Rev. Mol. Cell. Biol.* **2**, 369–377 (2001).
2. Hattori, M. & Minato, N. Rap1 GTPase: functions, regulation, and malignancy. *J. Biochem. (Tokyo)* **134**, 479–484 (2003).
3. Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299–1304 (2001).
4. Brinkmann, T. *et al.* Rap-specific GTPase activating protein follows an alternative mechanism. *J. Biol. Chem.* **277**, 12525–12531 (2002).
5. Rubinfeld, B. *et al.* Molecular cloning of a GTPase activating protein specific for the Krev-1 protein p21rap1. *Cell* **65**, 1033–1042 (1991).
6. Rubinfeld, B. *et al.* Localisation of the rap1GAP catalytic domain and sites of phosphorylation by mutational analysis. *Mol. Cell. Biol.* **12**, 4634–4642 (1992).
7. Hattori, M. *et al.* Molecular cloning of a novel mitogen-inducible nuclear protein with a Ran GTPase-activating domain that affects cell cycle progression. *Mol. Cell. Biol.* **15**, 552–560 (1995).
8. Gao, Q., Srinivasan, S., Boyer, S. N., Wazer, D. E. & Band, V. The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. *Mol. Cell. Biol.* **19**, 733–744 (1999).
9. Manning, B. D. & Cantley, L. C. Rheb fills a GAP between TSC and TOR. *Trends Biochem. Sci.* **28**, 573–576 (2003).
10. Gao, Q. *et al.* Human papillomavirus type 16 E6-induced degradation of E6TP1 correlates with its ability to immortalise human mammary epithelial cells. *J. Virol.* **75**, 4459–4466 (2001).
11. Ishida, D. *et al.* Myeloproliferative stem cell disorders by deregulated Rap1 activation in SPA-1-deficient mice. *Cancer Cell* **4**, 55–65 (2003).
12. Kraemer, A., Brinkmann, T., Plettner, I., Goody, R. & Wittinghofer, A. Fluorescently labelled guanine nucleotide binding proteins to analyse elementary steps of GAP-catalysed reactions. *J. Mol. Biol.* **324**, 763–774 (2002).
13. Chabre, M. Aluminofluoride and beryllofluoride complexes: a new phosphate analogs in enzymology. *Trends Biochem. Sci.* **15**, 6–10 (1990).
14. Scheffzek, K., Ahmadian, M. R. & Wittinghofer, A. GTPase-activating proteins: helping hands to complement an active site. *Trends Biochem. Sci.* **23**, 257–262 (1998).
15. Maheshwar, M. M. *et al.* The GAP-related domain of tuberlin, the product of the TSC2 gene, is a target for missense mutations in tuberous sclerosis. *Hum. Mol. Genet.* **6**, 1991–1996 (1997).
16. Jones, A. C. *et al.* Comprehensive mutation analysis of TSC1 and TSC2 and phenotypic correlations in 150 families with tuberous sclerosis. *Am. J. Hum. Genet.* **64**, 1305–1315 (1999).
17. Au, K. S. *et al.* Germ-line mutational analysis of the TSC2 gene in 90 tuberous-sclerosis patients. *Am. J. Hum. Genet.* **62**, 286–294 (1998).
18. Klose, A. *et al.* Selective inactivation of neurofibromin GAP activity in neurofibromatosis type 1. *Hum. Mol. Genet.* **7**, 1261–1268 (1998).
19. Xu, X., Wang, Y., Barry, D. C., Chanock, S. J. & Bokoch, G. M. Guanine nucleotide binding properties of Rac2 mutant proteins and analysis of the responsiveness to guanine nucleotide dissociation stimulator. *Biochemistry* **36**, 626–632 (1997).
20. Seewald, M. J., Körner, C., Wittinghofer, A. & Vetter, I. R. RanGAP mediates GTP hydrolysis without an arginine finger. *Nature* **415**, 662–666 (2002).
21. De Antoni, A., Schmitzova, J., Trepte, H. H., Gallwitz, D. & Albert, S. Significance of GTP hydrolysis in Ypt1p-regulated endoplasmic reticulum to Golgi transport revealed by the analysis of two novel Ypt1-GAPs. *J. Biol. Chem.* **277**, 41023–41031 (2002).
22. Ahmadian, M. R., Stege, P., Scheffzek, K. & Wittinghofer, A. Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nature Struct. Biol.* **4**, 686–689 (1997).
23. Albert, S., Will, E. & Gallwitz, D. Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J.* **18**, 5216–5225 (1999).
24. Nassar, N., Hoffman, G. R., Manor, D., Clardy, J. C. & Cerione, R. A. Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. *Nature Struct. Biol.* **5**, 1047–1052 (1998).
25. Graham, D. L., Eccleston, J. F., Chung, C. W. & Lowe, P. N. Magnesium fluoride-dependent binding of small G proteins to their GTPase-activating proteins. *Biochemistry* **38**, 14981–14987 (1999).
26. Goldberg, J. Structural and functional analysis of the ARF1-ARFGAP complex reveals a role for coatomer in GTP hydrolysis. *Cell* **96**, 893–902 (1999).
27. Scheffzek, K. *et al.* The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338 (1997).
28. Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gamblin, S. J. Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature* **389**, 758–762 (1997).
29. Daumke, O., Wittinghofer, A. & Weyand, M. Purification, crystallisation and preliminary structural characterisation of human Rap1GAP. *Acta Crystallogr. D* **60**, 752–754 (2004).
30. Lenzen, C., Cool, R. H. & Wittinghofer, A. Analysis of intrinsic and CDC25-stimulated guanine nucleotide exchange of p21ras-nucleotide complexes by fluorescence measurements. *Methods Enzymol.* **255**, 95–109 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Krämer for the gift of Rap1-Aedans-GTP. We thank I. Schlichting, W. Blankenfeldt, A. Scheidig, A. Rak, E. Wolf and O. Yildiz for data collection and crystallographic advice and the ESRF beam staff of beamline ID14-1 in Grenoble for support. O.D. thanks the Boehringer Ingelheim Fonds and P.P.C. the International Max-Planck Research School for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.W. (alfred.wittinghofer@mpi-dortmund.mpg.de). Coordinates have been deposited to the Protein Data Bank under accession code 1SRQ.

A conformational switch controls hepatitis delta virus ribozyme catalysis

Ailong Ke¹, Kaihong Zhou², Fang Ding¹, Jamie H. D. Cate¹ & Jennifer A. Doudna^{1,2}

¹Department of Molecular and Cell Biology and Department of Chemistry, and ²Howard Hughes Medical Institute, University of California at Berkeley, Berkeley, California 94705, USA

Ribozymes enhance chemical reaction rates using many of the same catalytic strategies as protein enzymes. In the hepatitis delta virus (HDV) ribozyme, site-specific self-cleavage of the viral RNA phosphodiester backbone^{1–3} requires both divalent cations and a cytidine nucleotide^{4–6}. General acid–base catalysis^{7–12}, substrate destabilization^{1,13} and global and local conformational changes^{14,15} have all been proposed to contribute to the ribozyme catalytic mechanism. Here we report ten crystal structures of the HDV ribozyme in its pre-cleaved state, showing that cytidine is positioned to activate the 2'-OH nucleophile in the precursor structure. This observation supports its proposed role as a general base in the reaction mechanism. Comparison of crystal structures of the ribozyme in the pre- and post-cleavage states reveals a significant conformational change in the RNA after cleavage and that a catalytically critical divalent metal ion from the active site is ejected. The HDV ribozyme has remarkable chemical similarity to protein ribonucleases and to zymogens for which conformational dynamics are integral to biological activity. This finding implies that RNA structural rearrangements control the reactivity of ribozymes and ribonucleoprotein enzymes.

To elucidate the catalytic mechanism used by the HDV ribozyme, we determined ten crystal structures of precursor forms of the ribozyme in which catalysis was prevented by mutation of the catalytically essential cytidine 75, chelation of divalent metal ions or 2'-deoxy substitution of the 2'-OH nucleophile (see Supplementary Table S1). Structures of the C75U mutant ribozyme and the wild-type genomic precursor ribozyme were solved at resolutions up to 2.2 Å and 3.4 Å, respectively. In all cases the structure of the precursor ribozyme resembles that of the product in overall conformation (Fig. 1a, b). The electron density at the active site of the C75U mutant reveals that the substrate RNA strand preceding the cleavage site makes a sharp ~180° turn about the scissile phosphate between the G1 and U –1 bases, packing the substrate strand into the P1 major groove and possibly destabilizing it in the ground state¹³ (Fig. 1c). Only the sugar–phosphate backbone density is visible preceding the scissile phosphate, and there appears to be no strong interaction between the substrate bases and the rest of the ribozyme, consistent with the low sequence preference 5' to the cleavage site^{16,17}.

In contrast to the structure of the product form of the HDV ribozyme⁷, electron density for a divalent cation coordinated to the 5' leaving oxygen is observed in the precursor ribozyme structures (Fig. 1c). Although the position of the active site metal ion varies slightly in the C75U mutant structures, in all cases it binds by outer-sphere coordination to six functional groups (four phosphoryl oxygen and two uracil keto oxygen atoms) that form a pocket adjacent to the scissile bond (Fig. 1d; Supplementary Fig. S1). Interestingly, the wild-type HDV ribozyme that was prepared and crystallized in the absence of divalent metal ions has the same structure as that observed for the C75U mutant precursor (r.m.s.d. of 0.30 Å for all-phosphorus-atom superposition), yet lacks any trace of the metal ion in the active-site metal-binding pocket

(Fig. 2a). This observation rules out disruption of active site geometry by the C75U mutation and shows that divalent metal ions are not required for HDV ribozyme folding, in contrast to some other ribozymes (reviewed in ref. 18).

To investigate whether the divalent cation observed in the active site participates directly in catalysis, we first showed that, consistent with previous results^{4,5}, the ribozyme is active in the presence of a wide spectrum of metal ions, including Mg^{2+} , Ca^{2+} , Sr^{2+} , Ba^{2+} , Mn^{2+} and Co^{2+} . It has trace activity in Cd^{2+} and Ni^{2+} solutions, no activity in the presence of Cu^{2+} and is specifically inhibited by $Co(NH_3)_6^{3+}$ (data not shown). To test the correlation between activity and metal-ion binding, various divalent cations including Mg^{2+} , Sr^{2+} , Ba^{2+} and Mn^{2+} were introduced into the C75U ribozyme crystals and shown to be present in the active site in

difference and anomalous difference electron-density maps (Supplementary Table S1). By contrast, Cu^{2+} , a cation that fails to support HDV ribozyme cleavage, does not bind in the active site. Furthermore, the inhibitory cations $Co(NH_3)_6^{3+}$ and $Ir(NH_3)_6^{3+}$ occupy the same site as the divalent ion and produce strong anomalous diffraction signals even after soaking crystals in 0.5–1 mM hexamine in the presence of 20 mM competing Mg^{2+} or Sr^{2+} . $Co(NH_3)_6^{3+}$ is also a competitive inhibitor of imidazole-catalysed cleavage of the C75U mutant ribozyme (Supplementary Fig. S2). Together, these data show a strict correlation between the activity of the ribozyme and the presence of a hydrated divalent cation at the active site. Importantly, a divalent ion binds similarly in the active site of a wild-type precursor ribozyme, as revealed in a difference electron-density map comparing structures of this ribo-

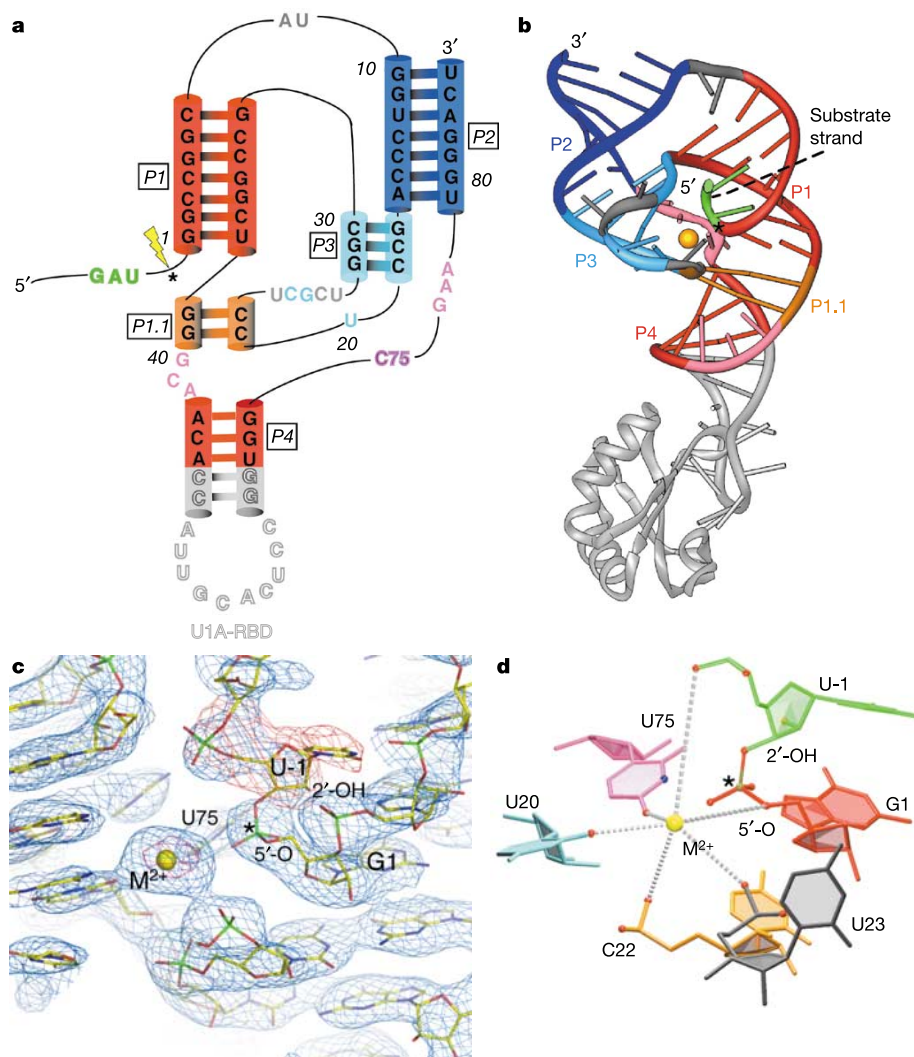


Figure 1 Precursor HDV ribozyme crystal structure. **a**, Sequence and secondary structure of the precursor-form HDV ribozyme crystallization constructs. Structural element colours match the three-dimensional representation in subsequent figures: substrate strand, green; P1, P4, red; J1/2, L3, grey; P2, blue; P3, light blue; J1.1/4, J4/2, pink; P1.1, orange; U1A-RBD and binding site, silver. Scissile phosphate is labelled with an asterisk and a yellow arrow. **b**, Ribbon-stick representation of the precursor HDV ribozyme, colour-coded as in **a**. Divalent metal ion, gold. **c**, Experimental electron-density map at the active site of the Sr^{2+} -bound C75U mutant precursor ribozyme, superimposed with the structural model. The 2.45 Å sigma A-weighted $2F_o - F_c$ map (contoured at 1σ) (blue) was calculated with the scissile phosphate and the -1 base omitted. An omit $F_o - F_c$ difference map (3σ , red) calculated from a 2.9 Å resolution $Ir(NH_3)_6^{3+}$ -bound C75U ribozyme structure reveals a 'predominant' position of the -1 base. The anomalous

difference density contoured at 6σ (magenta) around the Sr^{2+} ion is coincident with a ball-shaped density of 8σ in the $2F_o - F_c$ map. **d**, The active site configuration in the Sr^{2+} -bound C75U mutant precursor ribozyme. The Sr^{2+} is surrounded by the following ligands (with indicated coordination distances): O2 of U20 (3.2 Å), O4 of U75 (2.7 Å), phosphate of U23 (4.2 Å), O5' of G1 (5.2 Å), phosphate of C22 (5.1 Å) and phosphate of U-1 (5.2 Å). For reference, the distance between the metal and the 5'-O of G1 is 4.3 Å in the Mg^{2+} -bound structure, consistent with water-mediated contact. The distance from the metal to O4 of U75 is 4.0 Å in the $Co(NH_3)_6^{3+}$ -bound structure. The attacking 2'-OH of U-1 in its predominant conformation is 5.5 Å away from the N3 of U75. Occupancies of Mg^{2+} , Mn^{2+} , Sr^{2+} , Ba^{2+} , $Co(NH_3)_6^{3+}$ and $Ir(NH_3)_6^{3+}$ in the active site are ~0.5–1.0, whereas the anomalous difference signal of Cu^{2+} is not above background.

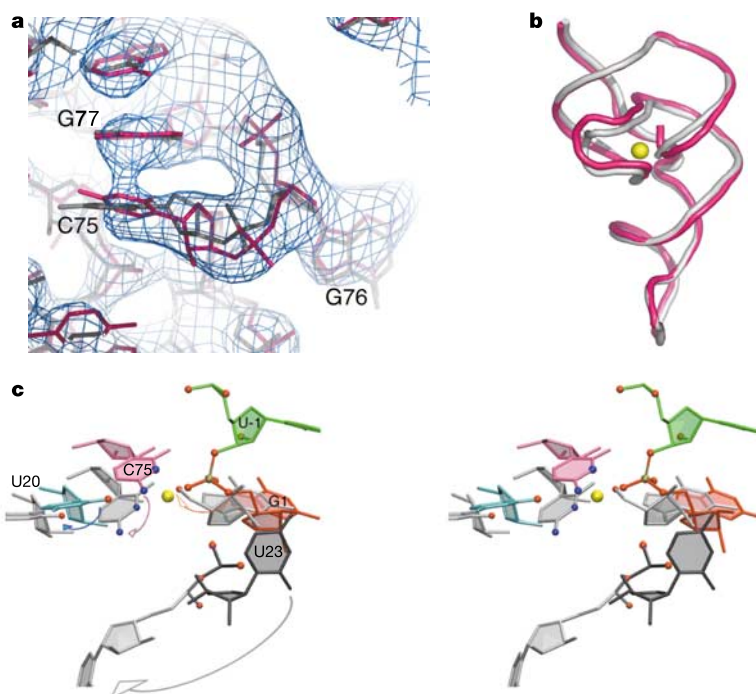


Figure 2 Conformational changes in the active site accompany HDV ribozyme cleavage. **a**, 3.4 Å experimental electron-density map of the wild-type precursor HDV ribozyme (1 σ) superimposed on the refined wild-type structure model (magenta) and that of the C75U mutant (grey). **b**, Backbone alignment of the precursor (magenta) and product (grey)

ribozyme structures. **c**, Stereo view of the aligned active sites of the precursor (coloured) and product (grey) ribozyme structures. Conformational changes were modelled using the C75U mutant ribozyme structure, shown by arrows.

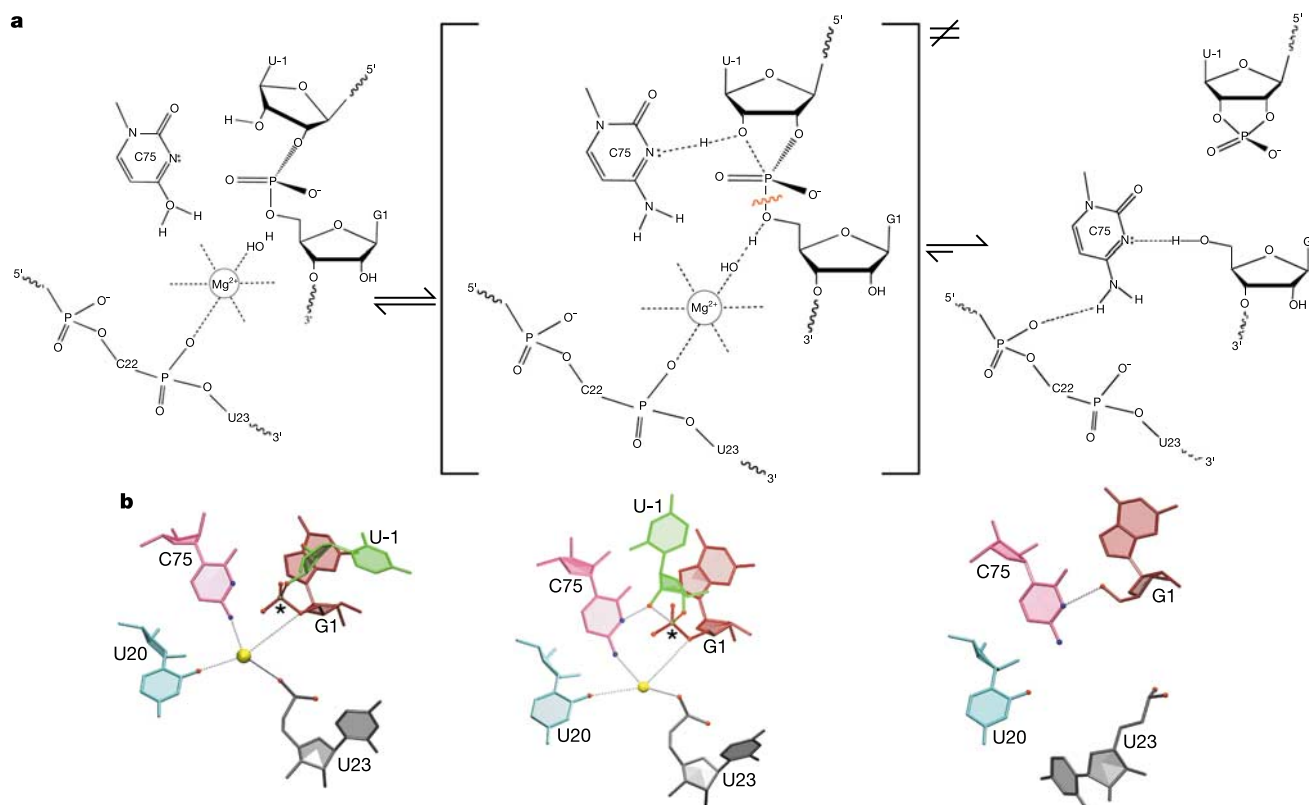


Figure 3 Proposed mechanism for general acid-base catalysis in the HDV ribozyme. **a**, In the ground state, a hinge rotation along O3'-P bond brings the nucleophilic 2'-OH to the in-line attack conformation. In the transition state, C75 (the general base) deprotonates the 2'-OH whereas the bound hydrated metal ion (the general acid) protonates the 5' oxygen leaving group. Conformational changes after scissile bond breakage discharge the

catalytic metal ion and down-shift the catalytic base C75 by 2 Å to enable hydrogen bonding with the 5'-OH of G1. **b**, Structural models of the HDV ribozyme in the ground state, transition state and product state. Some metal-chelating groups and coordinating waters are omitted for clarity.

zyme with and without divalent cations (Supplementary Fig. S1).

Strikingly, comparison of the precursor and product ribozyme structures reveals a significant local conformational change in the active site, resulting in dissociation of the catalytic metal ion after cleavage. Product-strand dissociation results in collapse of the P1 and P3 helices towards the centre of the ribozyme, pushing the ribose of the G1 base closer to C75 and deeper into the active site and closing off the binding site for the substrate strand. Removal of the scissile and the -1 phosphates upon product-strand dissociation triggers movement of the functional groups around the divalent cation (Fig. 2b, c). Furthermore, the C75 base shifts 2 Å deeper into the active site and forms a hydrogen bond with the 5'-OH leaving group.

It has been proposed that imidazole restores reactivity in HDV ribozymes with mutations at the active site C75 by supplying the missing general base functionality required to activate the 2'-OH nucleophile^{10,11}. Although no imidazole-binding sites were observed in electron-density maps determined using diffraction data from imidazole-soaked C75U mutant crystals, the crystalline ribozymes undergo imidazole-induced self-cleavage without disrupting the crystal diffraction quality. In addition, C75U mutant ribozyme RNA was crystallized following incubation in an imidazole buffer solution for several days, avoiding any potential crystal lattice artefacts. The structures of these 'product' forms of the C75U mutant ribozyme (Supplementary Table S1) show that although the conformational change partially occurs and the catalytic metal ion dissociates, the U75 base does not move into position for hydrogen bonding with the 5'-OH. The importance of this motion and the resulting structural collapse of the product active site is emphasized by the fact that a U analogue with a pK_a near neutrality does not rescue ribozyme activity when substituted for U75¹⁹. The essential nature of the exocyclic amine of C75 can be explained by its requirement for stabilizing the active site conformational change that controls catalysis. These data point to an interesting parallel between the HDV ribozyme and pro-region-containing enzymes such as caspases, in which conformational changes induced upon cleavage of the inactive zymogen form of the enzyme stabilize the catalytic cleft²⁰. Thus, in contrast to the hairpin and hammerhead ribozymes^{21–24}, the chemical step of the HDV ribozyme self-cleavage reaction is necessarily coupled to an active site rearrangement.

The precursor HDV ribozyme structures described above, together with previous biochemical data, are consistent with a model in which the sharply kinked RNA substrate is surrounded by two catalytic groups—a divalent metal ion coordinated through water to the 5' oxygen of the leaving group and the C75 base positioned near the 2'-OH of the -1 ribose. The pK_a of C75 may be shifted to near-neutrality by the closely positioned scissile phosphate, enabling it to accept a proton efficiently from the attacking 2'-OH nucleophile. Such a pK_a shift would only be observed in the precursor ribozyme, where the scissile phosphate is present, and is consistent with the lack of a significant pK_a shift in the product form of the ribozyme²⁵. The 2'-OH nucleophile in the ground-state conformation (Fig. 3, left) is 5.5 Å from the N3-H of U75 in the C75U mutant ribozyme structures and thus beyond hydrogen-bonding distance. It can be modelled within hydrogen-bonding distance of the U75 N3-H by a slight rigid-body rotation about the 3'-O-P bond of the -1 nucleotide. This rotated position brings the 2'-OH to the optimal in-line attack conformation (Fig. 3, middle), suggesting an additional role of C75 in the wild-type ribozyme in positioning the substrate. However, further rotation of U -1 to bring the 2'-OH near the metal ion results in steric clashes between the U -1 base and the P3 helix, leading us to conclude that the hydrated metal ion does not function as the general base in the reaction¹². Instead, C75 acts as a general base in the transition state to activate the 2'-OH, whereas the hydrated metal ion acts as a general acid to stabilize the 5' oxygen leaving group (Fig. 3). The proton transfer may occur in concerted steps in which a proton is

shuttled from the 2'-OH to C75, to $Mg(H_2O)$ through tautomerization, then to the 5' oxygen, consistent with observed pH-rate profiles that argue against formal loss of a proton by the hydrated metal ion¹¹. The conformational change in the active site following cleavage discharges the catalytic metal ion and displaces the C75 base such that the reverse (ligation) reaction is highly unfavourable (Fig. 3, right). Furthermore, closing of the space between the P1 and P3 helices blocks the binding site for the substrate strand. This explains why the HDV ribozyme is a dedicated nuclease and suggests that the virus evolved this unique catalyst to control the processing of replication intermediates during viral infection. □

Methods

RNA and protein preparation

Plasmids pL3 and pT3 encode the HDV wild-type and C75U mutant genomic ribozymes, respectively, with a U1A-binding site engineered in stem-loop P4', and a 3-nucleotide GAU sequence preceding the cleavage site. RNA was prepared by standard T7 polymerase run-off transcription using linearized plasmid and purified by denaturing gel electrophoresis. The wild-type precursor ribozyme RNA with a 2'-deoxy modification in the U -1 base was generated by DNA splint-mediated ligation²⁶. Native and selenomethionyl-derivatized U1A-RBD proteins were produced as described previously⁸.

Crystallization and diffraction data measurement

To obtain precursor ribozyme crystals, wild-type or C75U ribozyme RNA complexed with U1A-RBD protein at 0.5 mM concentration was diluted into nine volumes of denaturing buffer A containing 25 mM Tris-HCl pH 7.5, 20 mM NaCl, 2.5 mM $MgCl_2$ (replaced by 20 mM EDTA for the wild-type RNA) and 8 M urea. After overnight dialysis against this buffer, the complex was transferred in three 3-hour dialysis steps into buffer A containing 4 M, 2 M and 0 M urea, respectively, and reconcentrated to 0.5 mM. The wild-type ribozyme-U1A-RBD complex was crystallized by hanging drop vapour diffusion with a reservoir solution containing 5–10% (v/v) ($\pm$)-2-methyl-2,4-pentanediol (MPD), 50 mM sodium cacodylate pH 6.0, 40–80 mM NaCl, 20 mM EDTA and 5–15 mM spermine-HCl. The C75U mutant and the wild-type ribozyme with 2'-deoxy modification were crystallized in similar conditions or in the presence of 20 mM $MgCl_2$, $SrCl_2$ or $BaCl_2$ in place of EDTA. Structures of the C75U mutant ribozyme with Sr^{2+} or Ba^{2+} were determined from crystals grown in the presence of those ions. Mg^{2+} , Mn^{2+} , $Co(NH_3)_6^{3+}$, $Ir(NH_3)_6^{3+}$ and Cu^{2+} were introduced into the C75U ribozyme crystals by soaking in solutions containing 20 mM metal ion for 20 min, 30 min, 2 h, 6 h or overnight. The wild-type and mutant ribozyme RNAs in the crystals were confirmed to be in the precursor form by dissolving crystals after diffraction data collection and analysing the RNA on a 10% urea-polyacrylamide gel electrophoresis (PAGE). All crystals belonged to space group R32, with unit cell dimensions similar to those of the HDV product ribozyme crystals⁷. Native and anomalous diffraction data sets (near the absorption edge of the metal where possible) were measured for each of the metal-bound ribozyme crystal forms at beamlines 8.3.1, 8.2.1 and 8.2.2 at the Advanced Light Source (Supplementary Table S1) and processed using SCALEPACK2000²⁷.

Phase determination and structure refinement

The best phase information for the precursor ribozyme structure came from the Se-MAD (multi-wavelength anomalous diffraction) data set from a crystal of the C75U ribozyme mutant-selenomethionyl U1A-RBD complex. All four selenium sites were located using the program SOLVE²⁸. Heavy-atom parameters were further refined using CNS²⁹, followed by density modification and phase extension to 2.3 Å against the Sr^{2+} -containing native data set. One $Ir(NH_3)_6^{3+}$ site was found and refined in the Ir-SAD (single-wavelength anomalous diffraction) data from a C75U mutant crystal soaked overnight in 1 mM iridium hexamine using SOLVE. Se- and Ir-phased electron-density maps were used to build the precursor HDV ribozyme model. The precursor ribozyme model was then used to perform molecular replacement using AMoRe (CCP4 suite³⁰) on the rest of the native data sets. To determine the conformation of the cleaved C75U mutant ribozyme, the refolded C75U mutant ribozyme-U1A-RBD complex was dialysed against buffer A with 100 mM imidazole for 3 days; following this treatment, the crystallized complex was shown by urea-PAGE gel analysis to contain ~100% cleaved ribozyme. To test whether there were discrete imidazole-binding sites in the mutant ribozyme, crystals were soaked in buffer containing 100 mM imidazole for 1 h followed by diffraction data collection. No electron density corresponding to imidazole was observed in $2F_o - F_c$ and $F_o - F_c$ electron-density maps, consistent with the observation that the K_d of imidazole for the ribozyme is $>1.2 M^{10}$.

Refinement was carried out using CNS with 10% of the reflections removed for R_{free} factor calculation. Following rigid-body refinement, simulated annealing was performed on each refined structure to remove model bias. Ribozyme-bound metal ions were identified in anomalous difference electron-density maps. Alternating rounds of positional (Powell minimization), individual B-factor refinement and manual rebuilding were carried out until the R and R_{free} factors were below 30%. The sugar-phosphate density for the -1 base was clear at this stage for structures with resolution better than 2.6 Å. Together with the 'predominant' -1 base position density identified in the $Ir(NH_3)_6^{3+}$ -bound ribozyme structure, the entire -1 base could be built into the precursor ribozyme structure with confidence. Although residual density is present, the -2 and -3 bases were omitted from the models, except for the Sr^{2+} -bound structure that contains the

–2 base. TLS³⁰ and REFMAC5 were used to further refine the Sr²⁺-bound C75U mutant and wild-type ribozyme structures by modelling diffraction anisotropy.

Received 22 December 2003; accepted 23 March 2004; doi:10.1038/nature02522.

1. Shih, I. H. & Been, M. D. Catalytic strategies of the hepatitis delta virus ribozymes. *Annu. Rev. Biochem.* **71**, 887–917 (2002).
2. Rosenstein, S. P. & Been, M. D. Self-cleavage of hepatitis delta virus genomic strand RNA is enhanced under partially denaturing conditions. *Biochemistry* **29**, 8011–8016 (1990).
3. Reid, C. E. & Lazinski, D. W. A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proc. Natl Acad. Sci. USA* **97**, 424–429 (2000).
4. Suh, Y. A., Kumar, P. K., Taira, K. & Nishikawa, S. Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res.* **21**, 3277–3280 (1993).
5. Nakano, S., Proctor, D. J. & Bevilacqua, P. C. Mechanistic characterization of the HDV genomic ribozyme: assessing the catalytic and structural contributions of divalent metal ions within a multichannel reaction mechanism. *Biochemistry* **40**, 12022–12038 (2001).
6. Tanner, N. K. *et al.* A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr. Biol.* **4**, 488–498 (1994).
7. Ferre-D'Amare, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574 (1998).
8. Ferre-D'Amare, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556 (2000).
9. Oyelere, A. K., Kardon, J. R. & Strobel, S. A. pK(a) perturbation in genomic Hepatitis Delta Virus ribozyme catalysis evidenced by nucleotide analogue interference mapping. *Biochemistry* **41**, 3667–3675 (2002).
10. Perrotta, A. T., Shih, I. & Been, M. D. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, 123–126 (1999).
11. Shih, I. H. & Been, M. D. Involvement of a cytosine side chain in proton transfer in the rate-determining step of ribozyme self-cleavage. *Proc. Natl Acad. Sci. USA* **98**, 1489–1494 (2001).
12. Nakano, S., Chadalavada, D. M. & Bevilacqua, P. C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493–1497 (2000).
13. Shih, I. & Been, M. D. Energetic contribution of non-essential 5' sequence to catalysis in a hepatitis delta virus ribozyme. *EMBO J.* **20**, 4884–4891 (2001).
14. Pereira, M. J., Harris, D. A., Rueda, D. & Walter, N. G. Reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry* **41**, 730–740 (2002).
15. Harris, D. A., Rueda, D. & Walter, N. G. Local conformational changes in the catalytic core of the trans-acting hepatitis delta virus ribozyme accompany catalysis. *Biochemistry* **41**, 12051–12061 (2002).
16. Perrotta, A. T. & Been, M. D. A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* **350**, 434–436 (1991).
17. Jeong, S., Sefcikova, J., Tinsley, R. A., Rueda, D. & Walter, N. G. Trans-acting hepatitis delta virus ribozyme: catalytic core and global structure are dependent on the 5' substrate sequence. *Biochemistry* **42**, 7727–7740 (2003).
18. Pyle, A. M. Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* **7**, 679–690 (2002).
19. Oyelere, A. K. & Strobel, S. A. Site specific incorporation of 6-azauridine into the genomic HDV ribozyme active site. *Nucleosides Nucleotides Nucleic Acids* **20**, 1851–1858 (2001).
20. Chai, J. *et al.* Crystal structure of a procaspase-7 zymogen: mechanisms of activation and substrate binding. *Cell* **107**, 399–407 (2001).
21. Rupert, P. B. & Ferre-D'Amare, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780–786 (2001).
22. Rupert, P. B., Massey, A. P., Sigurdsson, S. T. & Ferre-D'Amare, A. R. Transition state stabilization by a catalytic RNA. *Science* **298**, 1421–1424 (2002).
23. Murray, J. B., Dunham, C. M. & Scott, W. G. A pH-dependent conformational change, rather than the chemical step, appears to be rate-limiting in the hammerhead ribozyme cleavage reaction. *J. Mol. Biol.* **315**, 121–130 (2002).
24. Dunham, C. M., Murray, J. B. & Scott, W. G. A helical twist-induced conformational switch activates cleavage in the hammerhead ribozyme. *J. Mol. Biol.* **332**, 327–336 (2003).
25. Luptak, A., Ferre-D'Amare, A. R., Zhou, K., Zilm, K. W. & Doudna, J. A. Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J. Am. Chem. Soc.* **123**, 8447–8452 (2001).
26. Moore, M. J. & Query, C. C. Joining of RNAs by splinted ligation. *Methods Enzymol.* **317**, 109–123 (2000).
27. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861 (1999).
29. Brunger, A. T. *et al.* Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Cryst. D.* **54**, 905–921 (1998).

30. Collaborative Computational Project, The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D.* **50**, 760–763 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Diffraction data for this study were measured with the guidance of C. Ralston, J. Holton and co-workers. We thank M. Been, K. Karbstein, J. Piccirilli, R. Spanggord and N. Walter for comments on the manuscript and for sharing unpublished data. This work was supported in part by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.D. (doudna@uclink.berkeley.edu). The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank with accession codes 1SJ3 (C75U/Mg²⁺), 1VBY (C75U/Mn²⁺), 1SJF (C75U/Co(NH₃)₆³⁺), 1SJ4 (C75U/Cu²⁺), 1VBX (C75U/EDTA), 1VBZ (C75U/Ba²⁺), 1VC0 (C75U/Sr²⁺/Imidazole), 1VC6 (C75U Cleaved Product), 1VC7 (C75U/Sr²⁺) and 1VC5 (Wild Type).

corrigendum

Characterization of a common precursor population for dendritic cells

Gloria Martínez del Hoyo, Pilar Martín, Héctor Hernández Vargas, Sara Ruiz, Cristina Fernández Arias & Carlos Ardavin

Nature **415**, 1043–1047 (2002).

In this Letter, we characterized a common precursor population for dendritic cells (DCs) isolated from mouse blood by their capacity to generate all DC subpopulations present in mouse lymphoid organs after transfer into irradiated recipients, including CD8⁺ and CD8⁺ DCs, as well as B220⁺ plasmacytoid DCs. Phenotypically, they were defined as CD11c⁺ MHC class II⁺ CD11b⁺ B220⁺ CD62L⁺. However, recent results from our laboratory (C.A., Beatriz León, Verónica Parrillas and G.M.d.H., unpublished work) indicate that, owing to a defect in the isolation method that was used, this common DC precursor population was highly contaminated by circulating NK cells, which express NK1.1 and DX5, display cytolytic activity and are devoid of DC-reconstitution potential when transferred into irradiated mice (C.A., unpublished results). Therefore, the cell population described in our original report comprised CD11c⁺ B220⁺ CD62L⁺ F4/80⁺ DX5⁺ NK cells and a population of CD11c⁺ B220⁺ CD62L⁺ F4/80⁺ DX5⁺ precursor cells with DC differentiation potential. Consequently, as the latter could potentially represent a heterogeneous population containing more than one DC precursor population, the proposed existence of the common DC precursors remains to be established. Experiments are underway to clarify this situation. □

Japan



Editor, Nature:
Philip Campbell

Publisher:
Peter Collins

Supplement Editor:
David Cyranoski

Subeditor:
James Kingsland

Production Editor:
Maria Hodges

Art Editor:
Martin Harrison

Layouts:
Vernon Adams

Picture Researcher:
Barbara Izdebska

Diagrams:
Charlotte Cartwright
Alisdair Macdonald

Nik Spencer

Production:
Yvonne Strong

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel 44 (0) 20 7833 4000
e-mail: nature@nature.com
http://www.nature.com/
nature



nature publishing group

© 2004 Nature Publishing Group

Japan is changing. The cynic might say that this can't be so — that Japanese policy-makers always talk about change but that nothing ever happens. It is indeed true that new policies in Japan often end up having little ultimate effect, that new systems are trumped by conventional ways of doing things. The latest policy initiative — the reorganization of the universities into administratively independent organizations — might seem to be just another in a long line.

With the university reform, will the system become more fluid and adaptable? Will the university presidents become strong, independent leaders? Will young researchers be able to follow their creative scientific wits? Will changes to the funding system force these things to happen?

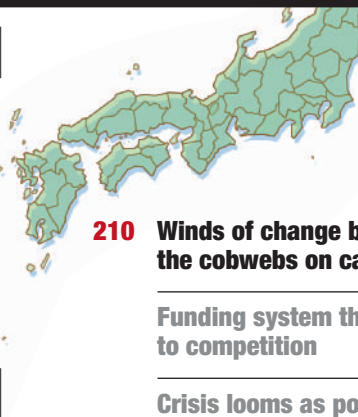
It is too early to judge the success of the many new policies described in the following pages. But several genuine indications of a changing Japan — both inside and outside the universities — are described here. University departments are launching term-based hiring and opening up positions for young researchers, university professors are testing the waters of industrial collaboration, and industrial researchers are taking their employers to court to gain recognition and recompense for their valuable research.

Such episodes illustrate a change in attitudes that will be given even greater momentum by the university reform. Until now, the pockets of change have been limited by a rather intractable framework — Japan's national university system. Now this framework itself is loosening. Whether or not the university reorganization actually achieves its goals, it is giving people room to think in new ways about how best to promote research. Researchers are asserting themselves more aggressively in university policy-making and in collaborations with industry. The expanding university engagement with industry could make a difference, especially for some of Japan's strongest suits, such as materials science and robot research, as well as biotechnology, in which Japan hopes to become internationally competitive.

And if, despite everything, the system still seems stultifying, Japanese researchers could leave it all behind and move to the island of Okinawa, where a new and innovative research institute is going up.

The coming year will tell us much about how these changes will play out, and whether they will eventually bear the desired fruit.

David Cyranoski



210 Winds of change blow away the cobwebs on campus

Funding system thrown open to competition

Crisis looms as population shrinks

Strange bedfellows sire hopeful monsters

216 Technology transfer: curiosity makes way for capitalism

A bridge to the rest of the world

Robots seek human inspiration

220 Will creativity thrive in an island paradise?



p210



p216

Winds of change blow away the cobwebs on campus

Japan's hidebound university system is being reformed.

Something has gone awry in peaceful Japan. In the past year, a researcher sued his university when his contract was not renewed, a university president was run out of office by faculty members, and, one after another, industrial scientists have sued former employers for not fairly rewarding them for products they helped develop.

Gone are the halcyon days when industrial and university employees accepted their lot in return for benefits such as the security of lifetime employment. A protracted recession has triggered a drive for change with restructuring efforts in recent years affecting most of the country's academic and industrial research laboratories.

The last, largest and most unwieldy piece in this restructuring puzzle is the reorganization of the university system. Several of the 99 national universities have already been merged, leaving 89. On 1 April, in the most comprehensive overhaul of the system since it was established after the Second World War, these survivors became semi-autonomous.

The new system removes university staff from civil-service rosters, thereby ending their guaranteed lifetime employment. In addition, universities must now undergo frequent evaluations to get money. Researchers' livelihoods will increasingly depend on winning competitive research grants. The upside is that they are now freer to collaborate with industry. At the same time, presidents have been given expanded powers and responsibility to make their campuses attractive to both students and academics — which will be vital in this era of increased competition.

Reformers hope this competitive boost will give the country's scientific stars, especially the young ones, room to flourish in a system that has been crowded out by unproductive professors. "Now universities must become the frontrunners of new knowledge, and everybody agrees on that," says Ken-ichi Arai, former director of the University of Tokyo's Institute of Medical Science and one of the loudest voices for change. "I never thought that would happen."

But Arai and many others worry that efforts to change the universities are not backed by changes in funding mechanisms. Many features of the reforms, such as expanding the role of presidents and breaking down hierarchical relationships between professors and their underlings, are bogged down in debate and cultural inertia.

Nevertheless, the focus on profit-making marks a definite shift for Japanese universities. Many researchers complain that the reorganization pushes science too far towards the sole pursuit of money in strategic fields set by the government. Universities are becoming ever more zealous in their efforts to transfer technology to industry. At the same time, researchers are becoming more vocal as they assert themselves within this flux.

New kids on the block

This change of attitude might be the most important outcome of the university reorganization, but it will also be difficult to guide. The new university presidents, granted unprecedented authority to determine their institutions' course, will have a big task on their hands. "Until now, the university president was largely a symbolic figure," says Takeshi Sasaki, president of the University of Tokyo. In the past, universities were run by de facto coalitions of departments, he says, each deriving its support directly from the government. "Each faculty protects its own interests," agrees Tadamitsu



Takeshi Sasaki: enhanced presidential powers?



Kishimoto, an immunologist and president of Osaka University until August 2003.

Farewell, then, to the levelling society created by department-based university funding, which provided university professors with stable levels of space, students and salaries. Stripped of their civil-servant status, researchers and staff will now have to perform to earn the salaries that let them eat and the grants that let them do science.

The new system will give presidents the power to take decisive action when a new field of science is needed to replace one past its prime, and to divert money to lure promising scientists. "We need presidents who can adjust the university's course quickly, even if this means that a department might have to be suddenly cut back," says Kishimoto, now a member of the Council for Science and Technology Policy (CSTP), the country's top science policy-making body.

Outside the universities, change is already beginning. "Freedom to follow new research fields has been the greatest benefit," says Hiroyuki Yoshikawa of his enhanced powers as president of the National Institute of Advanced Industrial Science and Technology (AIST), the research arm of the economics ministry. Since AIST became autonomous in 2000, he has pushed through plans to squeeze the institute's many materials and



Further funding reforms are needed to give young researchers more resources and freedom.

transgenic mice. “Young researchers in Japan can work independently on smaller projects, but they are not free to think of larger strategies,” he says.

The *koza* system is enshrined in a university law that sharply limits the activities of assistant and associate professors. In practice, junior faculty members do not even determine how their own grants are used. The science and technology ministry (MEXT) is considering changes to that law, but many researchers remain sceptical. The current reforms cannot touch the system, says Yoshihito Osada, vice-president of Hokkaido University. “No one can tell professors how to run their labs.”

The real problem is funding. Given the limited money available to young scientists, they have little choice but to take their place in this iron hierarchy. Grants for junior researchers, which generally do not include start-up costs for a new laboratory, are tiny. This forces them to spend all their time writing grant renewals, says Arai. Any grants that are sufficient to support an independent position go to senior researchers. In Japan, even more than elsewhere, the older you are the bigger your grant (see graph, below).

A major part of this problem is an evaluation system heavily tilted to past successes. “If Watson or Crick had been here they would have been kicked out before they were able to make their discoveries,” says Motoya Katsuki, director of the National Institute for Basic Biology in Okazaki. Indeed, a recent survey by the CSTP found that grants were ‘bunching up’, with the same ten researchers receiving nine or more of the country’s major

engineering laboratories into two. At the same time, AIST has pushed into newer fields such as bionics and biosensors. The products of this innovation, for example energy-producing materials studded with photo-receptor proteins taken from bacteria, can “answer industry needs,” says Yoshikawa.

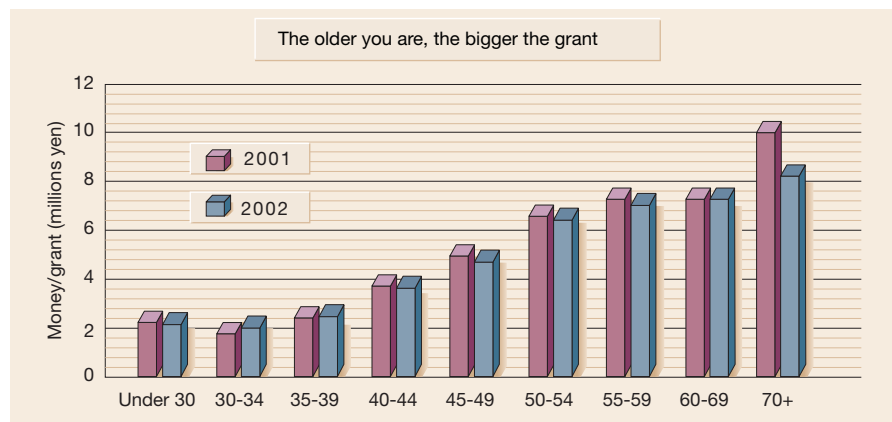
But the universities are a different species — it won’t be so easy for their leaders to get their way. Hiroaki Yanagida, a materials scientist and former president of Nagoya Institute of Technology, resigned in January after efforts to broaden his decision-making powers faced opposition. Sasaki, meanwhile, suspects his new powers might not extend beyond some ability to allocate extra resources as departments continue to determine their own fates. “No president will have power to scrap everything,” he says.

Some policy-makers argue that change must come from the outside. Foreigners will now be eligible to become presidents. Biologist and Nobel laureate Sydney Brenner, approached on a visit to California by Japan’s minister of science and technology, has agreed informally to head Japan’s latest experiment with university research institutions (see ‘Will creativity thrive in an island paradise?’ page 220).

Standing in the way of any true presidential authority are the *koza* (chair) professors

who dominate higher education. These formidable academics often run huge labs that visiting researchers working in more modest conditions have been known to call ‘factories’. Critics say the limits put on the freedom of graduate students, postdoctoral assistants and junior faculty members — researchers in their prime — take a tremendous toll on Japan’s overall creativity. And they cost the country some of its greatest talent.

Masashi Yanagisawa, now at of the University of Texas Southwestern in Dallas, was working at Kyoto University School of Medicine but quit Japan in 1991. He says he wanted more breathing space to work on



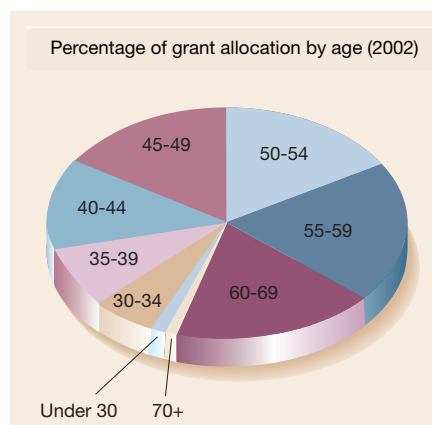
grants every year. Breaking into this realm will be no easy feat.

But reforms of the evaluation system might have to wait for deep changes in Japanese culture. “We are not used to evaluations,” says Yoshikawa. “If A gives B a bad review, B will hold a grudge. Evaluations aren’t used for encouragement of change — though this is starting to happen.”

Many large projects then turn out to be a waste, says Yanagisawa. He suggests single grants worth hundreds of millions of yen be broken down to make ten young researchers independent. “Because of the concentration of resources in these large laboratories, there must be ten times as many independent investigators in the United States as in Japan. That means ten times as many opportunities to hit on something really unique,” he says.

Some of the few meaty grants available to young researchers are Precursory Research for Embryonic Science and Technology (PRESTO) grants, but there is concern that these are on the wane. The Japan Science and Technology Corporation, which distributes them, denies this, but PRESTO grant distribution fell by more than 50% in 2003 as the government limited applications to fields such as quantum information processing.

The postdoctoral system, nonexistent until recently, has created opportunities but



also problems for young researchers. Japan has now met the target of 10,000 postdoctoral positions that the government set ten years ago. Despite their training, however, there are no university positions for the postdocs to fill — short of returning to a *koza*’s bottom rung. And Japanese industrial labs traditionally opt for employees fresh out of university whom they can train in-house. “The Japanese postdoc is not a gate to independence,” says Arai. “It’s another level of slavery.” Yamamoto says the glut of unemployed postdoctoral students is beginning to discourage potential academics from starting research careers. “If the job market doesn’t loosen up,

Masayuki Yamamoto welcomes evaluations.



DAVID CYRANOSKI

we’re going to have serious problems.”

Some attempts are being made. Shigetada Nakanishi and Tasuku Honjo, successive deans of Kyoto University’s faculty of medicine, started a programme to give younger scholars a chance to pursue their own projects. The five-year project, budgeted at about ¥900 million (US\$8.3 million) per year, recruited 22 researchers under 35 to study ageing, neurobiology and basic medical sciences. But when these researchers finish their five-year term they will still have to look for a niche in the rigid Japanese system, along with so many other young researchers.

Will the end of civil-servant employment

Funding system thrown open to competition

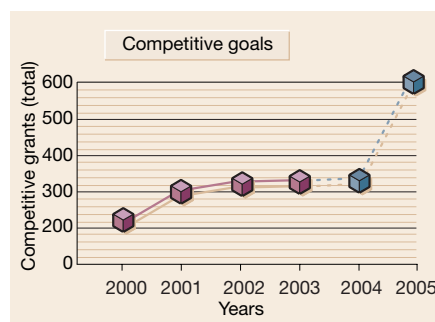
For years, Japan’s science policy-makers have advocated putting more money into competitive grants to produce more creative research. In response, scientists have called for an expanded evaluation system with more input from active researchers. Although expansion in competitive grants has been slow, the researchers’ demands may soon be met.

In 2001, the Council for Science and Technology Policy (CSTP), the country’s highest science policy-making body, called for competitive grants to double by 2005. With only a year left, competitive funds are increasing but at nothing near the pace needed to meet this goal (see graph, right).

The number of grant applications is expected to soar after April’s university reorganization, giving the funding agencies plenty to do. Scientists have previously served on advisory and review committees, but are now being hired

to work inside the agencies themselves.

Last July, the Japan Science and Technology Agency, one of the country’s foremost funding bodies, created the Research and Development Strategy Center, a think-tank of 28 researchers headed by Ryoji Noyori, a chemistry Nobel laureate and president of the Institute of Chemical and Physical Sciences in Wako. The



researchers will analyse trends and make policy recommendations to the education ministry and the CSTP.

The Japan Society for the Promotion of Science, which funds research outside the government’s strategic areas, also established a centre last July, bringing on board 50 scientists to participate in grant allocation decision-making. Their number is set to double this year. The Ministry of Education, Culture, Sports, Science and Technology also hired 25 young researchers this year to work on one of its major competitive grant programmes.

Scientists deserve a bigger voice in determining how the funds that keep them going are distributed, says Kaoru Okamoto, director of the Scientific Aid Division at the ministry. “Funding should be a user-friendly system, so we should ask the users — scientists — and not just bureaucrats.”

I.C.

Crisis looms as population shrinks

The once-packed lecture halls of Japanese universities are becoming lonelier and emptier.

Twenty years ago, a place at university was a hard-won ticket to a stable lifetime of employment. The obstacle to that good job for life was the gruelling entrance exam, and students crammed day in and day out for years in a bid to outperform the rest of the pack. But things are changing.

Decades of steady decline in Japan's birth rate (down from 2 per person in 1960 to 1.33 in 2001) have led to a corresponding decline in eligible young people to fill university places. Today there are 1.46 million 18-year-olds in Japan, down from a peak of more than 2 million in the early 1990s (see graph, below) and the figure is projected to fall to 800,000 by 2050.

Universities are faced with fewer and fewer applicants every year. In 2001, one quarter of Japanese universities had

more seats available than applicants and it has been estimated that in five years' time, there will be a seat somewhere for every student who wants one.

Japan's elite universities are just as competitive as ever, but the rest are resorting to more aggressive measures to fill their lecture halls. Universities now routinely entice students on TV and radio, even plastering the subways with adverts. They offer inducements in the form of scholarships for students with talents in areas other than academic subjects.

Even the dreaded entrance exams are getting easier, in part because students are not as well prepared as they used to be. Since 1998, with the never-ending studying and stress of 'exam hell' leaving schoolchildren in a state of exhaustion, the Japanese government has been reducing their burden by cutting back

elementary and high-school curricula, particularly in science. Akira Arimoto, director of the Research Institute for Higher Education at Hiroshima University, thinks this trend has gone too far. He says the amount of class time devoted to science has been cut so much in recent years that interest in science and mathematics is declining. More and more high-school students think "science and maths are just tools for passing entrance exams", he says.

Shinichi Yamamoto, director of the Research Center for University Studies at the University of Tsukuba, agrees. He notes ruefully that back in 1967 when he applied to university, even social-studies students had to take exams in several science subjects. "Now they have to take only one subject, or none at all."

A dwindling population of students combined with an exodus of the country's best and brightest young scientists overseas makes for a gloomy

picture of Japan's scientific prospects, says Arimoto. This changing demographic will eventually lead to a shrinking scientific labour force, part of a larger nationwide shortage that many observers say can only be filled through the immigration of thousands of foreign workers. But thus far Japan has not had much success recruiting workers from overseas.

In the immediate future, many universities will have to find ways to stay solvent despite empty lecture halls. Some of the original 99 national universities have already merged to make 89, and it is estimated that up to 200 private universities will be forced to close in the next ten years.

As competition between universities increases, education ministry officials hope that the ones left standing will be able to maintain Japan's high standards of research.

I.C.



REUTERS/CORBIS

status and the introduction of short-term contracts loosen up this system? Until recently, firing people was almost unheard of in Japan. Attempts by some university departments and research institutions to go short term suggest that it won't be easy. Recently Kazutomo Inoue, a developmental biologist at Kyoto University, sued his university claiming that it did not abide by the decision of the external evaluation committee in refusing to renew his contract.

Yoshikawa didn't even bother pushing for an end to the civil-servant hiring policies during AIST's reorganization programme. "If you cut anybody, they just sue," he says.

Some ambitious researchers, however, see short-term contracts as a golden opportunity. Masayuki Yamamoto of Tsukuba University's Center for Tsukuba Advanced Research Alliance, says he was glad to be one of the first in Japan to go on a term system when it was made legal to do so in 1998. "You can use the evaluations — assuming they're good — to make more demands on the university," he says. He has been happy to trade his promise of lifetime employment for more laboratory space.

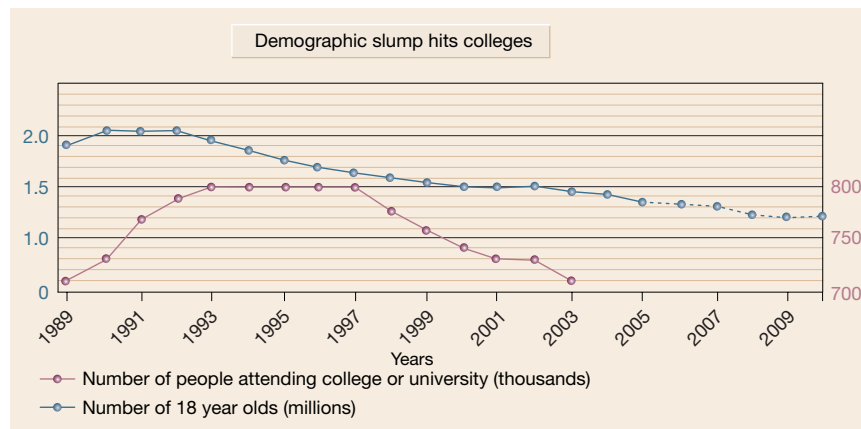
Outside academia, centres opened up by the Institutes of Physical and Chemical Sciences, which itself became an independent body last October, run on a system that is far from traditional. At the Brain Science Institute in Wako and the Center for Developmental Biology in Kobe, all lab heads are non-tenured, operate independently, and submit to regular outside evaluations for contract renewal.

Strapped for cash

Will the universities follow suit? Many remain sceptical. Arai fears that the current reforms will result in only superficial administrative reshuffling. "The title, cover, first page, may be changed. But after the third page, the real contents of the book may still be the same," he says.

Money problems might well force some changes. With little income from donations, universities are strapped for cash. In the future, they will be competing for a shrinking number of students (see 'Funding system thrown open to competition', opposite). Universities are responding with increased efforts to supplement their income through collaborations with industry.

And competition for state funding could



SOURCE: MEXT

Strange bedfellows sire hopeful monsters

Universities are not the only Japanese organizations getting squeezed. Cutting the number of civil servants and bureaucratic organizations has become a government obsession that often produces strange combinations. This is particularly true of the merger this April of the country's 14 'inter-university research institutions'. Researchers worry that this streamlining zeal might merely be adding onerous bureaucratic obstacles.

Seemingly random reconfigurations are nothing new to Japan. Following a streamlining effort in 2001 that cut the number of government ministers from 26 to 18, the minister in charge of science and technology took on responsibility for the distant island of Okinawa, and the disputed northern islands (the sticking point in failed negotiations that make Japan and Russia still technically at war).

But here, at least, this odd combination of roles did produce plans for a modern science and technology institute in Okinawa (see 'Will creativity thrive in an island paradise?' page 220). With luck, the same organizational magic will work on the education ministry's inter-university research institutes. These account for some of Japan's most innovative basic science in fields from particle physics to evolutionary genetics. Many, such as the High Energy Accelerator Research Organization and the National Astronomical Observatory of Japan, use costly, large devices that need government support.

Original plans to reduce their administrative load by merging them into one Max Planck-type institute (see *Nature* **415**, 352; 2002) met with resistance. A compromise plan, which came into effect on 1 April, divides the institutes into four groupings, some of which demand some

imagination to understand their underlying logic.

For example, the biology-focused Okazaki National Research Institutes will be lumped in with the National Astronomical Observatory and the National Institute for Fusion Science. How will the Okazaki institutes benefit from ties with the astronomical observatory, which runs massive projects like the Subaru telescope on Mauna Kea in Hawaii?

Some researchers at Okazaki, where small research projects produce many of Japan's most cited papers, are dubious. One molecular biologist there is worried about budget allocations and the additional layers of red tape involved in organizing a conference or other projects. "The way of doing things among the institutes is completely different. They spend money on big equipment and huge projects. We are much smaller." **D.C.**

get fierce. As part of the reorganization, the government said it would cut back its hand-outs and divert funds to competitive grants, doubling them over the five-year period from 2000 to 2005. Progress has been slow, however. Competitive funding has increased, but it is nowhere close to the target (see graph, page 212). And even with CSTP members such as Kishimoto favouring a 30% cut in university budgets, a powerful coalition of presidents blocked any cuts last year.

This has dismayed some central policy-makers who worry that short of funding changes, complacent professors will remain in their offices. "There's no place as comfortable as the Japanese universities — especially for professors without talent," says Masaaki Kimura, director of the Science and Technology Policy Planning division in the cabinet.

Even so, the CSTP has succeeded in using funding mechanisms to move research towards its strategic goals: information technology, nanotechnology and materials, life sciences and the environment.

Japan's largest strategic effort in the life sciences this year is a comprehensive analysis of the main mechanisms that initiate the expression of our genes, namely 'transcription factors' — proteins that latch on to the DNA strands — and promoter regions on the DNA itself that signal where transcription should begin. The project, centred on the Genome Sciences Center (GSC) in Yokohama, will have a massive ¥3-billion budget for each of the next five years.

It will begin by analysing some 1,500 transcription factors using 'transcription factor full-length cDNA' — isolated DNA molecules that each hold the genetic code for a given protein. Japan has the best collection of full-length cDNA for humans and mice, says Yoshiyuki Sakaki, who took over as GSC director on 1 April. "We want to take advantage of these," he says. The data will be made



It's getting ever easier to win a university place.

available to industry, which will use them to identify the essential metabolic 'switches' in cancer and diabetes.

The focus on strategic fields has worried some, especially when it is combined with short-term evaluations of progress. "It can take ten years of accumulation of seismological data to have a good result," says Toshitugu Fujii, director of the University of Tokyo's Volcano Research Center. Fujii says the larger universities have the resources to support such research, which falls outside of the government-designated strategic fields, but he worries about smaller institutions and the next generation of scientists. "They are going to turn towards fields that produce immediate results," he says.

Shunichi Amari, mathematician and director of the Brain Science Institute, agrees that basic science is suffering. "There is too much emphasis on technology that will have immediate impact with big money in the near future," he says.

Like it or not, the practical value of research is becoming a prominent concern for university and industry heads — and for researchers themselves. The biggest wake-up call came this January when a Tokyo court awarded ¥20 billion to the pioneer of blue

LEDs, Shuji Nakamura. In March, a materials researcher at Tohoku University followed in his footsteps, filing a suit against his former employer Toshiba claiming it did not fairly compensate him for his work on 'flash memory devices'.

Profit awareness is seeping into the universities as licensing of technologies and joint research projects with industry take off. University professors are also venturing into commerce with their own companies. The university system is opening up to encourage these advances (see 'Curiosity makes way for capitalism', page 216).

Building on tradition

Balancing the demands of basic research and education with the pursuit of industrial applications will not be easy. But the outlook is by no means grim. Japan is an international leader in fields such as materials and robotics, and the government provides hefty support every year for promising new areas such as measuring devices and regenerative medicine. Despite lean times for Japan's economy, the government is keeping faith in science as the engine for Japan's intellectual and social growth, putting ever more money into the research budget.

As Asia's scientific powerhouse, Japan outperforms startups such as China by far and looks set to continue to do so. The customary way of doing things might be under attack, but Japan's long tradition of research is what will keep the country ahead, says Muming Poo, a neuroscientist at the University of California, Berkeley. "It takes several generations to build up the tradition of doing first-rate science," he says. "In Japan there's a 150-year tradition of doing excellent science that didn't even stop during the war." ■

David Cyranoski is *Nature's* Asian-Pacific correspondent.
I-han Chou is assistant editor at *Nature Neuroscience*.

Curiosity makes way for capitalism

Universities can now profit from their creativity, but not all academics are pleased.

In December 2003, a company called OncoTherapy Science hit the Tokyo Stock Exchange at full tilt. The company, a spin-off from the University of Tokyo that provides potential targets for cancer therapy based on genetic analyses of patients, is now valued at about ¥110 billion (US\$1 billion). Established only three years ago, the company is already making a profit.

Its performance would have been impressive for a start-up anywhere in the world. In Japan, hampered by regulations and short on know-how for transferring knowledge from the university to industry, it was an extraordinary achievement.

So much so, say OncoTherapy executives, that it has become the target of envy. Yusuke Nakamura of the University of Tokyo, whose extensive research into the genetic profiles of cancer patients laid the foundations for the company, refuses to talk to journalists lest his mission in creating the company be misunderstood.

The situation highlights a tension within academic circles and among science policy-makers in Japan. If members of the government's central policy-making body, the Council for Science and Technology Policy, had their way, similar companies would be popping up all over the place. University professors sold on 'curiosity-driven' research, however, are often not interested and even resent the pressure to make their science practical. Some object to any university professor making a profit from publicly funded research.

To make matters worse, the government has often sent researchers mixed signals, encouraging them to seek industrial involvement but crimping their ability to do so. Not



any more. A new law on intellectual property has recently been passed to make it possible for universities to profit from their creativity and the message to researchers is loud and clear: make money.

Rushing to market

Last July, for example, the science and education ministry authorized 34 intellectual-property offices at universities throughout Japan. The economic ministry has set up 36 of its own technology-licensing offices to flag up and market any promising research results. Kobe's Translational Research Informatics Center is trying to provide a national framework for the clinical trials system, filling in the gaps in the process of taking university medical research through clinical trials and commercialization.

Japan's patent office, often criticized for its slow work and lack of expertise, has set the ambitious goal of becoming the world's fastest at processing applications. Over the next five years it will increase its workforce by nearly 50% to 1,600. In the past the office could not boast a single doctorate in the natural sciences, but there will be more than 20 in the new 98-strong class starting this month. These multidisciplinary

positions are becoming a career opportunity for Japan's glut of postdoctoral students.

Universities across the country are responding with administrative innovations. In snowy Sapporo, for example, the local government and industries have teamed up with Hokkaido University to create a 30-hectare business park housing academic and industry research centres in a bid to invigorate Hokkaido's stalled economy.

The park's headquarters, located in the university's Creative Research Initiative 'Sousei' (CRIS) unit, selects junior faculty members from different university departments. The scientists are excused from all teaching for up to seven years to focus exclusively on their research. Projects range from using DNA microarrays to monitor marine animal responses to pollution, to the creation of crystals that become electrically conductive when exposed to light.

The man at the helm is Hiroshi Takahashi, a former manager at Mitsubishi Chemical who now holds a professorship free of any teaching or research duties. "Research and education were sufficient roles for universities in the twentieth century," he says, "but in the twenty-first we have to think about how to use that knowledge."

An agreement signed last April with electronics heavyweight Hitachi launched

"University professors sold on curiosity-driven research resent the pressure to make their science practical."



collaborative tissue engineering and bio-informatics projects that started in October. And in January, Mitsubishi Heavy Industry signed up to a collaborative project to study new plant-based fuel sources and environmental monitoring via satellite.

Meanwhile, government initiatives have already started to pay dividends nationwide through large increases in industry collaborations. Universities are also producing many more patents and spin-off companies (see graph, page 218). Even in biotechnology, a field in which few companies actually get off the ground, nine university spin-offs are now listed on the Tokyo Stock Exchange.

Birth of the shrewd professor

The organizational changes that gave the universities their independence in April will accelerate this trend as Japan moves towards a US-style intellectual-property system. In 1999, a law was passed allowing universities to keep control of patents arising from government-funded research. The law was inspired by the US Bayh-Dole Act of 1980, which many credit with the birth of fruitful university-industry interaction in the United States. However, since the national universities in Japan were at that time government institutions, control of the patents reverted back to the government,

A bridge to the rest of the world

In January, Mauricio Terrones, a Mexican physicist, arrived at Japan's National Institute for Materials Science (NIMS) in Tsukuba. There he met an Indian chemist called Vinu Ajayan and a Russian materials scientist called Dmitri Golberg, who share an interest in carbon materials with nano-sized pores.

Over prolonged 'tea breaks' during Terrones' two-and-a-half-month stay, they transformed boron materials, some that Ajayan had developed in Japan and some that Terrones had brought from Mexico, into completely novel structures. Terrones (pictured below) says the new class of material will be the basis for at least three papers and could have several practical applications ranging from fuel cells to biosensors, in which it would serve as an anchor for thousands of proteins. "It's a new material with a lot of surface area," says Terrones.

Such international collaborations may be commonplace in somewhere like California, but projects forged with such speed and flair are something that Japan has long sought but rarely achieved. The relatively few foreigners who dare to come to Japan are often hampered by red tape and other obstacles.

This group's research was made possible by a new organization within the institute called the International Center for Young Scientists. Yoshio Bando, the materials scientist who created the programme, has recruited some 20 researchers from around the world who have completed their

doctorate within the past 10 years, provided them with generous funding (a ¥7-million salary and ¥5-million research grant for full-time fellows), and then let them devise their own research plans within fields ranging from medical applications for biocompatible materials to the nanoscale design of electronic devices.

Many of Japan's attempts to internationalize have used large grants to entice researchers but then left them isolated in large departments. Language presents a particular challenge. "If you're a postdoc and the only foreigner, it can be quite difficult," says Terrones. But just over a year into this five-year project, the centre already boasts fellows from 15 countries. Only one of the researchers is Japanese. All documents and meetings are in English. "We were able to get a critical mass of foreigners so they have no choice but to speak English," says Bando.

Bando hopes the project will be renewed when its five years are up. Either way, it could have a lasting impact. He plans to recruit ten full-time NIMS researchers from the centre. Those like Terrones who only stay for a short time will help to foster lasting international collaborations.

The centre will also be a model for universities and other institutions pursuing the elusive goals of international collaboration and fostering young scientific talent. "If we do it well, many will follow us," says Bando.

D.C.



making them difficult to offer to industry.

"Use of the patents was greatly restricted and complicated," says Koichi Sumikura, an intellectual-property expert at the National Graduate Institute for Policy Studies in Tokyo. This all changed in April. "From now on, professors will be able to make the most of their patents," says Sumikura.

And the pot is getting sweeter as universities start to compete to attract top-quality researchers. A recent survey by the *Nihon Keizai Shimbun* newspaper showed that universities were on average raising the amount that a professor could earn from a patent to,

on average, 30% of royalties or licensing fees. Kyushu Institute of Technology will give an attractive 56 per cent. "This is a remote place but we still want to attract the best people in the world," explains an official in the university's intellectual-property office.

If the experience of other Japanese research organizations is anything to go by, independence for the universities will quickly spur researchers' eagerness for commercial opportunities. The National Institute for Materials Science, which became independent in April 2001, doubled the number of patents it holds over the following

year. At the National Institute of Advanced Industrial Science and Technology, within a year of becoming independent, the number of industrial collaborations jumped from 5 to 78. Looser regulations were one factor, but the biggest difference was the researchers themselves, says Hiroyuki Yoshikawa, the industrial institute's director. "Their consciousness has changed."

Japan's technology-transfer offices are preparing for a possible upsurge. Last month, the Center for Advanced Science and Technology Incubation was appointed as the University of Tokyo's official technology-transfer outlet and is building its staff from 12 to 17. "We're getting busy," says Takafumi Yamamoto, the centre's president.

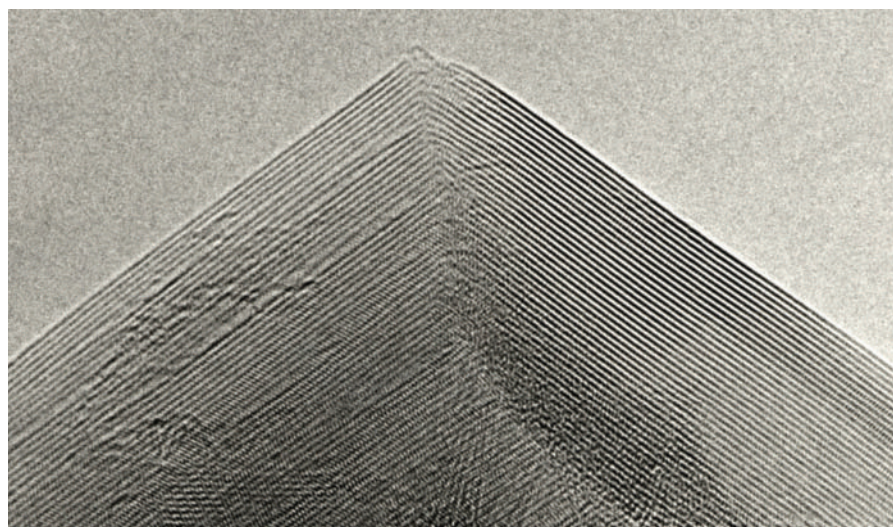
The greater awareness might overcome some of the obstacles that have made industry — especially overseas companies — shy away from collaboration with Japanese universities in the past. "Many Japanese researchers don't understand disclosure agreements, and foreign companies are worried about leakage," says Sumikura. "Professors need to learn for example to release data only after publication."

From the perspective of industry, the greater awareness will come as a mixed blessing, says Toshio Baba, manager of the nanotechnology group at NEC's Fundamental and Environmental Research Laboratories, which already takes part in many university collaborations. "University professors will definitely have a better business sense about what needs to be done to develop a project," he says. "It should be easier to work together." But, he adds, "Negotiations over patents will also be more difficult. In the past, universities often just gave away their intellectual property."

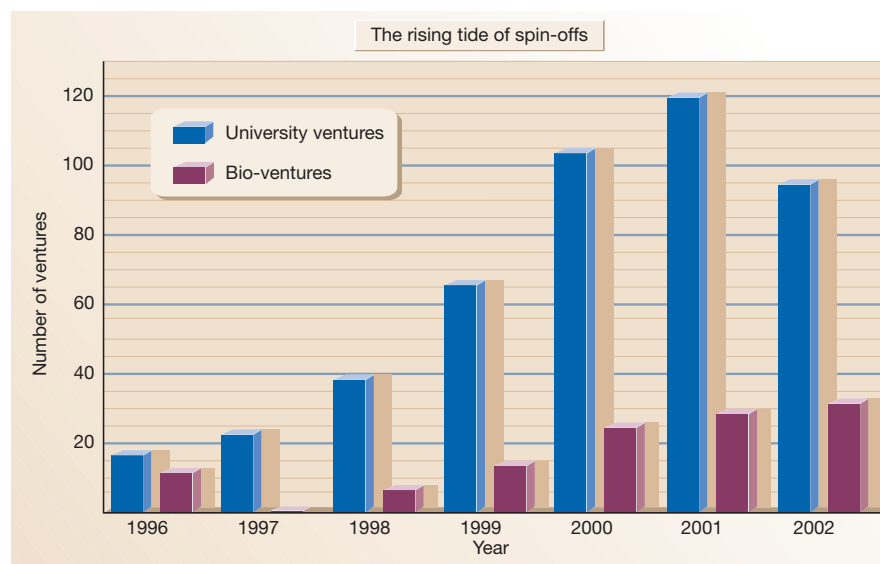
Cutting out the middle men

The system has some flaws. "I am not completely optimistic," says Sumikura. For one thing, Japan is lacking in multidisciplinary experts who combine legal and scientific training with business experience. "These resources that combine knowledge of business, science and law — we need them now. But it's going to take time."

Education initiatives, including a planned expansion of law schools to take in 5,590 new students, aim to cultivate specialists in intellectual property. These positions may also be a career opportunity for Japan's many unemployed postdoctoral students.



'Nanocone' of boron nitride produced in Dmitri Golberg's lab.



Robots seek human inspiration

Hopes are high that in the near future, robots will be personable humanoids that can walk the dog, work as salespeople and make good conversation. Japanese companies such as Honda, Sony and Toyota are locked in a fierce contest to turn out the best and brightest humanoid robot. These little corporate ambassadors strut their stuff for the cameras, shake hands with foreign dignitaries, inaugurate bullet trains and even conduct orchestras.

Many agree that the impressively stable and smooth gait of robots such as Honda's ASIMO are a good demonstration of how far robotic technology has advanced. But they also point out that robots are still a long way from being able to navigate unpredictable conditions and terrains. "Even the best conventional robots only operate under very controlled situations," says Yasuo Kuniyoshi, a roboticist at the University of Tokyo. "They fail if there is even a small perturbation." The solution he and a handful of other researchers advocate is to look to the best model of adaptability and intelligence at hand — humans.

One of the leaders of that movement is Mitsuo Kawato, director of Computational Neuroscience Laboratories at Advanced Telecommunications Research Institute International. For the past 15 years, Kawato has been exploring computational theories of brain function and testing them on robots. Using the types of learning used by animals, such as reinforcement, Kawato's group has taught their life-sized humanoid, DB (short for Dynamic Brain, pictured), to carry out more than 30 tasks, including juggling and playing air hockey.

Integrating neuroscience and robotics opens up new horizons for both fields, says Kawato. "Conventional neuroscience techniques are hopeless for understanding information processing, and robotics without biology is just engineering," he says.

Minoru Asada of Osaka University agrees. Asada studies how a robot with artificial vocal cords can learn to speak by interacting with a human caregiver. Kuniyoshi of the University of

Tokyo also examines adaptive imitative behaviour, but his research focuses on how robots solve problems like standing up from a prone position by extracting important principles about their own actions.

"These guys are on the frontier of brain science and robotics," says Hiroaki Kitano, director of Sony Computer Science Laboratories and the Kitano Symbiotic Systems Project. "Other people build robots for performance, but they are using robots as a vehicle to understand human intelligence."

Such approaches have caught the attention of companies in Japan. Jun Tani of the Brain Science

Institute is collaborating with Masato Ito of Sony Labs to explore how novel behaviours emerge when Sony's robot QRIO learns to mimic human behaviour. Sony is also collaborating with Kawato's group to develop walking technology.

And humanoid robots are capturing the public's imagination. "Outside Japan, people tend to think of humanoid robots as evil artefacts like Frankenstein's monster," says Kawato. There is no such problem in Japan. "It's the Japanese dream to have a robot with intellectual capabilities and a warm heart," says Tani. **I.C.**



Teaching new robots old tricks.

But will Japan ever be able to catch up with the rest of the world when it comes to converting academic vigour into commercial success? Education-ministry data show that in 2001 there were only 105 spin-offs from universities in Japan compared with 368 in the United States. And it is unlikely that foreign companies will rush to Japanese campuses the way companies around the world move into US campuses such as Stanford and MIT. "Now there are more opposite cases — Japanese companies collaborate with universities overseas," says Sumikura.

But with the new system in place, it might become easier for overseas companies to establish in Japan. Then it will be up to the universities to prove themselves. Researchers in some of the country's traditionally strong fields such as materials science and in budding ones such as regenerative medicine might well jump at the chance.

The Berlin-based pharmaceutical company Schering has bought a stake in Japan's future with a research institute that accounts for more than 5% of its global research budget. When it starts operation in July, there will be more than 40 scientists working on cell-therapy treatments for nervous system disorders and heart disease. In the future, the institute might also investigate regenerative medicine for treating cancer.

Government enthusiasm for regenerative medicine — especially in institutions such as the neighbouring Center for Developmental Biology — was one factor in convincing Schering to set up in the country, says Nikolaus Mueller, managing director and head of R&D in Japan. But even more important were the changes in the Japanese university research system that offered a channel to the likes of Tokyo, Kyoto and Kobe universities. "Professors and university hospitals have been very open-minded," he says. "We see that the doors are getting wider."

David Cyranoski is Nature's Asian-Pacific correspondent. Additional reporting by I-han Chou, assistant editor at Nature Neuroscience.

Web links

Japan Patent Office

♦ www.jpo.go.jp

METI technology licensing offices

♦ www.meti.go.jp/policy/innovation_corp/top-page.htm

Translational Research Informatics Center, Kobe

♦ www.tri-kobe.org

Will creativity thrive in an island paradise?

Okinawa is the unlikely setting for an ambitious institute.

A remote island best known for its US military base, scuba diving and goya (an intensely bitter melon) might not seem the obvious place to build a cutting-edge research institute from scratch. But the Japanese government has been pouring cash — ¥3.3 billion (US\$30 million) over the past two years — into a new institute to be based in a resort village called Onna on the island's west coast. The research community in Japan, including those who are involved with the project, is largely sceptical, but success here could mean a new model for an ailing university system.

For the cabinet minister who has been put in charge of not only science and technology but also Okinawan affairs, the Okinawa Institute of Science and Technology was a marriage of convenience. But it is set to grow into a fully fledged research university, and is a beacon of success for current efforts to fix the country's core university system.

Advocates of the research institute say its distance from Japan's administrative centres, which has stifled other universities, might be its greatest asset. "Okinawa has an open, southern culture, separate from Japan in both its workforce and philosophy," says Ken-ichi Arai, director of the Tokyo Metropolitan Institute of Medical Science. Arai hopes the university will form a bridge — geographically and culturally — with researchers based in other countries, such as Singapore.

This different way of thinking is reflected in the unprecedented selection of an overseas scientist as the institute's president. The minister of science and technology made personal visits to Nobel laureate Sydney Brenner to entice him to take the top post. Brenner, who says his appointment hasn't been formalized, admits there is a long road ahead: "They have had little experience with modern educational institutions."

If all goes according to plan, the institute

will use English as its primary language and maintain close connections with scientists in other countries. Half of its researchers will come from overseas. Unlike universities in the traditional Japanese mould, it will undertake numerous collaborations with industry and foster the careers of researchers young and old in a merit-based system.

But given the distance from major cities and lack of infrastructure, who will come?

Round one of recruitment has gone well — at least on the surface. In February, four precursor projects, which will be conducted in temporary lodgings near Onna until the institute's 2007 opening, were selected. The four project heads are all leaders in their respective fields.

Akira Tonomura, currently a researcher at Hitachi in Hatoyama near Tokyo and a

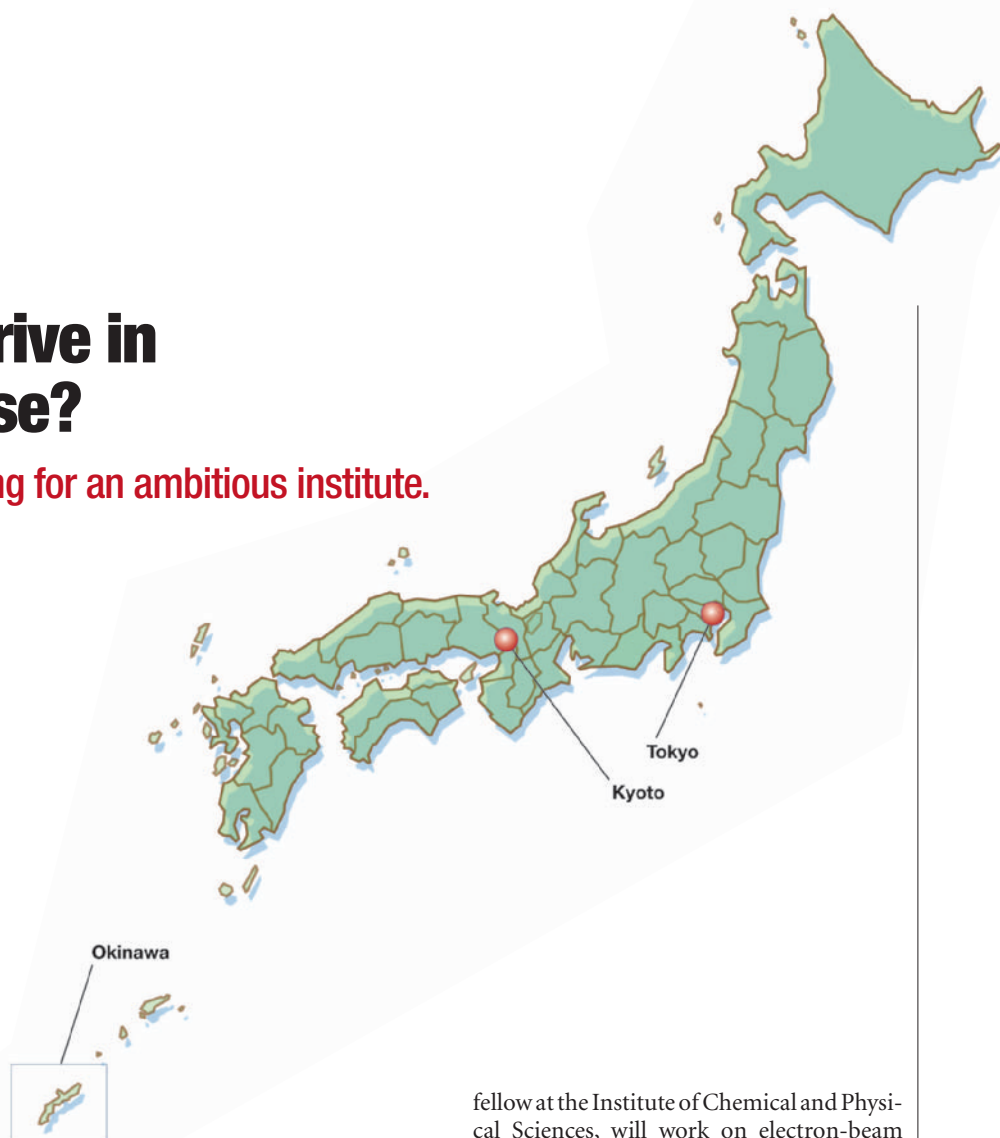
fellow at the Institute of Chemical and Physical Sciences, will work on electron-beam microscopy; Mitsuhiro Yanagida from Kyoto University will head a group looking into genes related to cell growth in yeast; Kenji Doya of the Advanced Telecommunications Research Institute in Kyoto will take a computational look into the 'molecular mechanisms of the mind'; and Shogo Endo of the Brain Science Institute in Wako near Tokyo will study the biological basis of learning and memory.

Excitement about the opportunity to do original science is mixed with some doubt over whether the institute will be able to draw

the world's best scientists, and whether the government will be able to maintain the high levels of funding.

Tonomura's project, for example, has already run into difficulties. Tonomura, who is known for his electron-beam observations of vortices of

magnetic flux in superconductors, plans to install a 300-kilovolt microscope in Okinawa. But he worries that the ground is not stable enough for his sensitive equipment.



"Excitement about the opportunity to do original science is mixed with doubt over whether the institute will draw the world's best scientists."

US military planes buzzing by and high-voltage electric lines could also disturb his measurements.

The site is being examined to see what preventive measures, such as an electromagnetic shield, might be put in place. The project launch, scheduled for last month, has been postponed, possibly for a year.

Tonomura also dreams of opening up completely new fields of science in Okinawa. The microscope he has in mind could focus its electron beam on individual atoms, localized electromagnetic fields and quantum phenomena with a resolution below the angstrom level. With such a microscope, says Tonomura, researchers could view the individual atoms of a carbon nanotube, for example.

The project would require a high-pressure electron microscope with improved lenses. "What has been holding us back is lack of effort," he says. There is one big hitch: the microscope would cost a cool ¥5 billion and take five years to develop.

Yanagida is also looking forward to a chance to do something completely new, but says the plans for Okinawa are still a bit "foggy". Best known for his work on chromosome segregation in fission yeast, his group will be looking at how this organism makes the decision about whether to lay low when nutrients are not present or divide when they are. Figuring out the genetic components of this decision will help us understand how, for example, liver cells continue to proliferate throughout adult life whereas brain cells do not, says Yanagida. "Yeast is a wonderful system to work on this problem," he says.

He continues to head a laboratory with 25 researchers at Kyoto University, where his project grants last for another two years and cannot be transferred to Okinawa. For the time being, he will straddle the two, visiting his group in Okinawa once a month.

Despite the ambiguity of his position, Yanagida hopes he can help make Okinawa a success. "We must prove that research can be done there," he says. "Then young people will be encouraged to go."

Other issues must be resolved. Conflicting opinions about whether to focus on basic research or develop spin-off technologies coexist uneasily in the institute's planning.

The determination to make it a bona fide international research institute is clear.



Sydney Brenner (left) may head the Okinawa institute, and Akira Tonomura leads one of its projects.



The Okinawa Institute of Science and Technology will share the island with a large US military base.

Yanagida says he was tickled by the opportunity to write his first grant application in English at the age of 62. "It was a good experience." But most remain dubious about the institute's ability to recruit top international scientists. Even Japanese scientists "need courage to go there", says Yanagida.

Doya is more optimistic. A leader in the study of human behavioural learning, he will shift his focus to the study of animals — rats — for the first time. He wants to discover how adaptive learning is implanted in the brain by dopamine when animals perform behaviours such as finding water or a mate. He hopes to transfer this knowledge

back to robots. "We need to know what kind of arrangements are necessary for this kind of flexible learning," he says.

He looks forward to a thriving scientific community developing on Okinawa. And coming from Kansai Science City in Kyoto, which has developed rapidly in the ten years since its establishment, he has no doubt that the infrastructure in Okinawa will evolve.

And if other scientists share his enthusiasm for windsurfing and scuba diving, there should be no trouble filling the vacancies at the new institute. "It's an ideal place for me," he says.

David Cyranoski is *Nature's* Asian-Pacific correspondent.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayourni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Aerospace meets biotech

In California, scores of laid-off aerospace workers are looking for jobs. Meanwhile, life-science companies in the region are canvassing for entry-level employees. The needs of both these unlikely bedfellows have been met by a state-sponsored programme that has, in effect, played matchmaker.

The demand for technical employees in areas such as biomanufacturing has increased. This is especially true in California, where many biotech companies are becoming more product-oriented, and so need people with manufacturing and quality-control skills, says Caz Pereira, a Bay Area consultant who specializes in California workforce issues. Such companies tend to be successful at attracting young scientists into their technical ranks — but they are often unable to keep them. Many youngsters do stints in industry to fill the gap between undergraduate and PhD work, to get an introduction to the industry before pursuing a business angle, or as a means to pay off student loans before doing graduate work. So there are often vacancies at the lower levels that need filling.

This insatiable need is now being met by the \$1.5-million programme, which is retraining workers to move into the biotech sector. The participating companies are free to interview and select the programme participants as they see fit, but they are encouraged by a partial wage subsidy to employ them. The programme aims to have trained and placed more than 400 workers in the life-sciences sector by the end of 2005.

The programme has broad implications. Other local companies are already enquiring about the scheme, and it comes at a time when Britain is investigating ways to help prepare its undergraduates for industry careers. Starting similar programmes in other countries might be one way to combat an anticipated skills shortage in both academia and industry.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

What's brewing in
food science?

p224

CAREER VIEW

Scientists & Societies

Young genomicists get
connected

Graduate Journal

Future echoes

Movers

Paul Cataford

p226

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

The yeast is rising

Makers of beer, wine and cheese need microbiologists to keep fermented products at their peak. Kendall Powell gets a taste of the career offerings.

Every student who has ever crossed paths with *Saccharomyces cerevisiae*, the academic version of brewer's yeast, has pondered getting a job at the local brewery. Cody Reif is living the dream. As a microbiologist at the New Belgium Brewing Company in Fort Collins, Colorado, Reif and his colleagues start each day by tasting the latest batches of the company's eight brews. He says that the group's taste-test gives better statistics for determining if a batch is up to company standards. Yeah, sure, statistics.

"Is working here as much fun as people think? Yes, it is," Reif says. The daily sampling illustrates a criterion for working with fermented food products not usually highlighted in a basic microbiology degree: critical sensory skills. It reflects the fact that the bottom line is not the science behind making fermented products, but rather the flavour and appeal of a product.

Careers focusing on the microbiological processes that go into making beer, wine and dairy products demand solid backgrounds in microbiology and biochemistry, but they also incorporate special expertise in food science and sensory training. Research projects in this small but diverse field range from troubleshooting contaminants that might spoil a vat of wine to designing probiotic microbial cultures for 'functional foods' that aim to improve the immune system.

An understanding of how fermenting microbes behave, and how their metabolites change the properties of foods and beverages, can be picked up in specialist degree programmes or through on-the-job

experience. Reif, who has a microbiology degree, brags: "I can probably identify four or five different strains of yeast or bacteria just by smell."

One place a microbiologist might seek work is in a microbrewery — a small regional enterprise. There, says Reif, it is necessary to know the ins and outs of the entire beer-making process because any microbial problems could bleed into the next phase. Also, scientists at smaller companies handle most analyses on site, from quality control and yeast propagation to monitoring for 'critical beer spoilers' — bugs that can crash the keg party.

This holds true at smaller wineries, too, says Lars

Bjorkman, an enologist at Flora Springs Winery in St Helena, California. His bachelor's degree in viticulture and enology from the University of California, Davis, combines classic microbiology, biochemistry and food science with sensory classes for the tailored enology half of the degree. These allow him to analyse his winery's varieties through each stage of production. Bjorkman says that a large part of his job is to ask each day, "What is growing in this wine?", and to find out if it is advantageous or a troublemaker.



Tread carefully: winemakers are looking for experienced microbiologists.

A STRANGE BREW

Larger enterprises such as Miller Brewing Company in Milwaukee, Wisconsin, and E. & J. Gallo Winery in Modesto, California, have their own research and development departments that employ scientists at bachelor's, master's and PhD levels. Tom Pugh, director of enology research at Gallo, says that the company does both applied research, to optimize fermentation conditions, and basic research that strives to identify the microbial characteristics that affect a wine's feel, flavour and aroma.

His research group looks at the way genetic differences between strains affect metabolism throughout the fermentation. Gallo also takes on microbiologists with degrees slanted towards wine-making as winemakers, he says, overseeing a wine's production from grape to bottle.

Similarly, David Ryder, chief brewmaster at Miller, says that microbiologists may want to become brewmasters at individual breweries. There's also a small group of microbiologists doing R&D at the corporate headquarters in Milwaukee. These, says Ryder, track yeast physiology and use DNA

Gut feeling for health

Developing probiotics — microbial cultures added to foods to convey an added health benefit, such as improved immune function — is one of the biggest growth areas in food science research. Probiotic cultures of bacteria normally found in foods are thought to colonize the gut and aid both digestion and immunity. But companies are still looking for the definitive study that links their consumption to good health. Experts say the most marketable background for work on probiotics would be a microbiology degree coupled with courses in human nutrition and immunology or with experience of running clinical trials.

K.P.

OBREMSKI/CORBIS



Say cheese: an understanding of fermentation — and how microbes affect taste and texture — is crucial in the dairy industry.

microarrays to find out if the yeast strains are ‘happy’. Happy yeast are healthy yeast that grow fast in a variety of environments and efficiently convert sugars to alcohol. Like Pugh, Ryder stresses that he looks for employees who are passionate about both the science and the product.

CHEESE WHIZZES

And what goes better after a fine wine than a nice hard cheese to cleanse the palate? Makers of cheese (and yoghurt) face an equally complex, though different, fermentation process. Microbiologists working on the dairy side of the fence also hunt for the microbes that bring the desired texture, flavour and aroma to fermented milk. Understanding how microbes will react under the different stresses of the cheese-making process is a key component not necessarily gleaned from working in an academic microbiology lab. This is where the practical side of a food-science degree comes in, says Kayla Polzin, a senior scientist and microbiologist at the Land O’Lakes dairy company’s labs in Arden Hills, Minnesota.

Polzin says that, in her job, a deep understanding of microbiology and metabolism is just as important as a broad understanding of the many microbes that can crop up in dairy cultures. She is one of two PhD-level microbiologists at Land O’Lakes, but many more dairy-food microbiologists are employed by the companies that provide the starter cultures. As well as developing cultures that are resistant to bacteriophages and viruses, or that work more efficiently, these scientists track down microbe characteristics responsible for specific flavours and textures sought by their customers.

“Pizza companies can be very, very demanding as to what the cheese looks like on their pizza. It has to brown, stretch and melt just the right way,” says Dennis

Romero, a molecular microbiologist at the starter-culture company Rhodia in Madison, Wisconsin. He is one of about two dozen research scientists at Rhodia, but he notes that most of its competitors are based in Europe and are ten times as large.

Romero stresses that researchers at starter-culture companies must do science that is practical at all times. The best way for microbiology students to gain this applied viewpoint of research, he says, is to work on an industry-funded collaboration. Starter-culture companies tend to ‘outsource’ the most basic research by setting up joint projects with an academic lab.

Christian Hansen, a food-ingredient company based in Hoersholm, Denmark, employs about 80 scientists in its central research laboratories. Chief science officer Peter Olesen says that — because the company deals in colours, flavours and enzymes as well as cultures — it employs protein chemists, chemical engineers and scientists with medical and veterinary degrees.

One of their recent basic microbiology research projects developed what Olesen calls a “revolution in lactic acid bacteria” — a strain engineered so that it can grow under aerobic conditions (with oxygen), resulting in cultures that ferment faster.

All areas of food microbiology value knowledge of food processes and fermentations more than a narrow focus on molecular mechanisms of organisms. For that reason, most of the opportunities in this field do not require postdoctoral training. Industry-related experience is rated highly — whether an internship, working on a wine harvest or an industrial collaboration.

Reif says that, all beer-drinking benefits aside, he gets job satisfaction from the critical thinking required in troubleshooting: “When problems come up it’s an opportunity to use your head.” ■

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

GRADUATE JOURNAL

Future echoes

Although I have at least a year left until I finish my PhD, going to the graduation parties of my friends has got me reflecting on my future. The happy faces and reminiscent moods of my colleagues make me think about what life might hold after my thesis.

The biggest decision for me and my colleagues is choosing between academia and industry. Universities have their attractions, but the grey zone between completing a PhD and becoming a full professor makes me hesitate. I believe that these years of uncertainty turn a lot of excellent people away from academia.

I feel that, if there are risks in the early stages of an academic career, I might as well take an even bigger chance and try to start a company. If that doesn't work out, I could consider a job with an established biotech or life-sciences firm. I could even see myself working in a totally different subject — as long as the people are nice and I find the work challenging and exciting.

I might not be tied to an academic future, but it nevertheless dominates my present. Before I think too far ahead I need to work on my analysis and get more content for my thesis. ■

Philipp Angerer is a second-year PhD student in biotechnology at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

SCIENTISTS & SOCIETIES

Young genomicists get connected

PhD students and postdoctoral fellows around the world struggle with the same everyday problems. What technique shall I use? How do I analyse these data? Where do I find this article? Is there a course on that subject? And their advisers are often too busy to provide an immediate answer. Arguably, life would be much easier for students and postdocs if they could share their questions, insights, new ideas or problems with scientists in the same career phase. After all, their colleagues are likely to have the same problems — and maybe even the same frustrations.

In the Netherlands, the Genomics Network for Young Scientists (GeNeYouS) was set up to make such interactions happen. As a result, young researchers from all over

the country and from all parts of genomics research gather both in person and on the Internet.

The Netherlands is a small country but there is genomics research under way at more than ten universities and a range of institutes — all of which have their own specialties and expertise. As the greatest distance between any two academic institutions takes about four hours to travel, it is possible to meet people from within the network regularly.

But for everyday business, a four-hour drive is still too long. Then the GeNeYouS website is the place to be. It provides a quick reference to courses on different topics given in different institutions, an up-to-date overview of career possibilities, a calendar with GeNeYouS and genomics events, and a forum to reach out to other young researchers. It also offers a search engine for jobs.

Currently the network has more than 330 members. The website is the backbone of the network, with all information about activities freely available. In addition, the network sends out a monthly digital newsletter to all of its members.

This January, GeNeYouS held its first symposium. Lectures and workshops led by experts in the field helped the 120 young researchers gathered there to make some tough everyday decisions: how to set up a genomics experiment, how to make your academic career successful, how to get funding. Once again, young researchers from all over the Netherlands struggled with some everyday problems. But this time they did it together. ■

Simon Mooijaart is a PhD student at the Leiden University Medical Center, the Netherlands, and is the founding chairman of GeNeYouS.

♦ www.geneyous.nl

MOVERS Paul Cataford, president, University Technologies International, University of Calgary, Canada



When the venture-capital market dried up a few years ago, Toronto-based investment banker Paul Cataford considered returning to university. "I was actually thinking about going back to school and doing some PhD research on technology transfer," he says. Instead, Cataford last month found himself at the helm of University Technologies International (UTI), the University of Calgary's technology-transfer office.

Researching the position was a valuable lesson in itself, Cataford says. When he got the initial job offer, he

examined the activities of several technology-transfer offices at Canadian universities. He found that few had launched successful spin-off companies — especially compared with the United States or Britain.

After more digging, Cataford decided that most Canadian universities had technology-transfer offices that were held closely by the university. UTI, on the other hand, was a wholly owned subsidiary but was kept "at arms length" by Calgary.

As a result, Cataford decided that UTI spin-offs would be more malleable to the influence of a technology-transfer office, just as start-ups can be shaped more by early-stage venture-capital firms than by later entries.

The switch from private to public wasn't the first drastic career change that Cataford has undergone. After first completing an engineering degree, he went into sales and marketing of computer equipment. But his penchant

for business led him to an MBA and, from there, to a finance career starting in investment banking, moving into private equity and ending up in venture capital.

Cataford realizes that the UTI job will be different from his earlier venture positions. He will have a much broader list of 'constituents' than simply investors and funded companies, including the university's faculty, its economic development board, other UTI employees and potential US partners.

And Cataford knows he faces a challenge in trying to develop a technology cluster when the region is small compared with, say, the US Bay Area. But Cataford is confident that by specializing in engineering, medical devices and global positioning technologies — all areas of expertise for Calgary — and by finding partners in the United States, UTI's arms-length relationship with its parent university will enable him to succeed. ■

CV

2002–04: Founding management partner, HorizonOne Asset Management in Toronto, Canada

2001–02: Executive managing director, BMO Nesbitt Burns Equity Partners in Toronto

1996–2001: Managing director and president, BCE Capital in Toronto